

Whipping up panic about the Pill

The British Committee on the Safety of Medicines seems to have acted with needless haste by warning women against forms of contraceptive pills whose risks may not be all that great.

LAST week was a week of high drama on the contraceptive front in Britain. Last Wednesday (18 October), the Department of Health sent a letter to all physicians in general practice, telling them that specified brands of contraceptive pills carry a risk of thrombosis that is twice as high as that caused by earlier contraceptive pills. Some doctors received this letter, and one of them, married to a journalist, told her husband, who promptly made a story of it. Other physicians had to wait longer for word of what the scare was about. Pill-using women, on the other hand, did not wait to besiege their physicians by telephone and otherwise. Then the general confusion was made worse by the arrival in person of Professor Walter Spitzer, of McGill University and the Potsdam Institute in Berlin, which appears to be in the business of designing and conducting large epidemiological studies. Spitzer, whose study of 6,000 women over five years had been cited by the British government, denounced both the Department of Health and its advisory body, the Committee on the Safety of Medicines (CSM) for hasty and premature action, causing an "epidemic of anxiety". All this came out at a press conference hastily arranged at London Airport.

The circumstances are more than a little odd. The forms of contraceptive pills now under a cloud contain synthetic analogues of female hormones that are more effective (gram for gram) than the constituents of earlier contraceptive pills. That was chiefly the reason why they were supposed to be less likely to cause thrombosis in those taking them over long periods. But since 1989 there have been reports from Germany that even the low-dose pill entailed the risk of thrombosis among those who used it. That, no doubt, is part of the reason why Schering commissioned Spitzer's five-year study.

To be fair, the CSM had other reasons to be concerned. In July, the World Health Organization produced preliminary results of an inquiry in 17 countries suggesting a link between the new pills and venous thrombosis. There is also an international study of doctors' records based on Boston University, while the CSM also says it had access to a "final" version of Spitzer's data. (Spitzer says that, nevertheless, it should have waited at least until the data had been submitted for publication and, in the process, peer reviewed.) All three studies, the CSM says, pointed in the same direction, that the risk of venous thrombosis with the new pill is roughly twice that with the now-safer ver-

sions of the old, and is roughly six times greater than the risk in women who take no pills at all. So, the CSM says, it had a moral duty to recommend quick action.

Really? The absolute numbers are as important as the relative risks. In Britain, roughly 25 per cent of women between the ages of 16 and 49 use contraceptive pills, perhaps half of them the newer versions, implying for a population of 58 million that there would be 1.5 million women at risk. If, as quoted, the risk of thrombosis with the new pills is less than one in 200,000 per year, the incidence of thrombosis would be of the order of one every 50 days among women taking the new pills. Granted that the CSM must be alert to the dangers in the use of the medicines licensed on its recommendations, the numbers hardly justify the haste with which the CSM appears to have made its recommendations on this occasion. Even pregnancy is not risk-free, after all.

The lessons to be learned from this curious tale are complicated, but are likely to be more than ever needed in the years ahead. Sudden announcements that widely used medicines have unexpected side-effects are almost always likely to cause panic, as on this occasion. The urgency with which bad news is made public must be commensurate with the absolute, not the relative, risk, uncertain though both may be. In this case, it would have been acceptable to let Spitzer's paper embark on the process of peer review. It might even have been found that the increased risk of thrombosis is offset by a decreased risk of other conditions — the 'old' pill is, for example, now known to protect against ovarian cancer. Who (except Spitzer) can know at this stage whether the new pills offer protection of a similar kind?

What this implies is that the CSM should have waited with its recommendation. A few weeks either way would have made no significant difference. That way, it would even have been able to devise some means of making the health department's blunt announcement intelligible to those affected. It is an assault on the supposed rationality of the communication between the regulators of social practices and their practitioners that the regulators should issue warnings without what is called, in the trade, counselling. And, whatever may be thought of the motives of the pharmaceutical companies, it is crass to take them by surprise when they are best equipped to put out health warnings to their ultimate customers and those people's physicians. □



'Millions of pounds' confirmed lost to universities over funding switch

London. The British government is bracing itself for the publication of an independent report that is expected to confirm that leading universities have lost many millions of pounds as a result of changes in the procedures by which they are reimbursed for some of the indirect costs of research.

The losses stem from a decision by the government to transfer responsibility for funding the overhead costs of research projects from the then Universities Funding Council to the five (now six) research councils. As a result, money for overheads previously paid as part of a block payment to universities, and then distributed to academic departments, is now linked directly to individual grants.

The purpose of the transfer has been to increase accountability for the use of such funds by ensuring that payments are linked more directly to the costs actually incurred by specific research projects. Indeed, at the time of the transfer, the government promised that the changes would be "fiscally neutral", and thus not affect either the overall funding of research or the total reimbursed to universities to cover indirect costs.

But a report being drawn up for the Office of Science and Technology (OST) by the management consultants Coopers

and Lybrand, following complaints from a number of leading universities, is expected to back their charges that the implementation of the changes has not worked out as planned.

As a result, says the report, a considerable sum — said by some to be as much as £15 million over the past few years — of the money transferred to the research councils has not made its way back to universities in the way the government had promised. Rather, the transferred money has been used to pay the direct costs of an increased number of research grants.

Although a draft of the Coopers and Lybrand report was completed several months ago, a final version has yet to be agreed with the OST. This has led to speculation that some of its conclusions may be watered down before they eventually appear in print. Nevertheless some of those who have seen the draft say the implications of the findings are embarrassing for the government.

According to Peter North, for example, vice-chancellor of the University of Oxford, the report confirms his university's view that the transfer of funds from the funding councils to the research councils has been "little short of disastrous" for all the major research universities.

"It has at last been recognized that the problem is what we have always said it was, namely that the same amount of funds is coming to universities, but in the form of a larger number of grants, each of which tends to be inadequately funded," said North in his annual address to the university senate earlier this month. "The consequence is that there is a real danger that the more grants a university gets, the less well-off it becomes. We can only hope and urge that that mistake will be corrected."

North's complaints are echoed among a small group of leading research institutions — including in particular the University of Cambridge, Imperial College London and University College London — which have already expressed their concern privately to government ministers on several occasions about the extra costs they are having to find from their own resources.

"We are as concerned as Oxford is," says Cyril Doherty, academic planning officer at Cambridge, which claims that last year it had to find an extra £3.5 million out of its own funds — compared to a total income from the research councils of £45.8 million — to cover items that would previously have been paid for out of their government block grant.

A crucial question raised by the Cooper and Lybrand report — and said to have contributed to the delay in publication — is who is responsible for the apparent inaccuracies in the government's initial estimates of the amounts of money that the universities would be awarded by the research councils after the lump-sum transfer.

Some in the universities blame the research councils for being excessively restrictive in applying the rules under which indirect costs can be claimed. "Some of the items that we thought were allowed under the new rules have been struck out, and others have been reduced," says Doherty.

But many research council officials, as well as administrators from other universities that claim not to have lost out on the transfer, argue that the shortcomings may lie with the research universities themselves for failing to give adequate attention to the detailed justifications now required for each item of indirect expenditure.

Either way, no major changes are expected as a result of the report — partly because the research councils are unlikely to be willing to see any of the transferred money returned to the four funding councils through which the government now supports universities in England, Wales, Scotland and Northern Ireland respectively.

David Dickson

Parliamentary panel to survive?

London. The future of the House of Commons select committee on science and technology, under threat earlier this year as a result of a reorganization of government portfolios (see *Nature* **376**, 103; 1995), seems assured following a statement of support from Ian Lang, the President of the Board of Trade, and as such cabinet minister responsible for science.

The statement was made during the annual debate in the House on UK science policy, in which Lang also claimed that, despite the charges of its critics, the Conservative government is committed to maintaining a healthy science base. "There will be no lurch to short-termism," he promised.

But if the committee — whose future lies formally in the hands of Parliament

itself — does survive, it is unlikely to make life comfortable for the government, particularly over the controversial decision which led to the threat of its own demise, namely to transfer the Office of Science and Technology to the Department of Trade and Industry.

There was widespread criticism of this decision, widely seen as removing science from its previous central position in the government's concerns, from members of the opposition Labour party, as well as various Conservatives.

Those on the government benches who revealed their unease at the move included Sir Giles Shaw, the chairman of the science and technology committee, which is responsible for overseeing all areas covered by the OST.

Earlier, in his introductory speech to the debate, Lang announced that the government has agreed to provide additional resources to support projects aimed at investigating the public understanding of science and technology.

D.D.



Shaw: 'anxious'

Canadian government set to announce new technology programme

Washington. Canada is about to launch a new technology development programme that will allow the government to lend around C\$300 million (US\$225 million) a year to industrial companies to support the development of high-technology products.

The National Technology Investment Programme, known as N-TIP, is expected to be the centrepiece of the Canadian government's long-delayed science and technology policy review (see *Nature* 376, 376; 1995), now due to be released in the middle of next month.

The review may also lead to the creation of a small secretariat at the Privy Council Office — the Canadian government office that supports the cabinet — to coordinate spending on science and technology across the government, as well as an outside board to advise the secretariat.

N-TIP is effectively a successor to the long-established Defence Industries Productivity Programme (DIPP), which is being phased out after funding of C\$140 million this year. The aerospace industry has been lobbying strongly for a replacement programme, particularly the aero-engine manufacturer Pratt & Whitney. But it will also be open to grant applications from other high-technology, export-led industrial sectors.

According to industry lobbyists, the money allocated to N-TIP will not be explicitly allocated to sectors by the government, but awarded to the strongest proposals. That will leave the aerospace companies well placed, given their previous experience with grant applications with DIPP, one lobbyist said.

Universities will be able to become involved in N-TIP through partnerships with industry. But they are still likely to look upon the programme as evidence that the Liberal government gives higher priority to the support of technology than to basic scientific research.

N-TIP grants will be repayable under a royalty scheme if the product developed is successful and profitable. The Canadian cabinet has agreed to the scheme in principle, and annual funding is expected to be between \$C200 million and \$C300 million, tending towards the higher figure if the scheme is extended to include Atomic Energy of Canada's work on environmental technology.

The government hopes that N-TIP will eventually become self-sustaining, on the basis of royalty income received from successful projects. But initial money for the programme will have to come from the existing budgets of the government departments involved.

C.M.

Clinton portrays science as the path to racial harmony

Washington. President Bill Clinton has attacked Republican plans to cut science and technology spending, arguing that such spending can "make real the promise of the American dream" and thus ultimately help soothe the racial tensions now racking the United States.

In a 20-minute address — his first on science in three years in office — delivered at the White House last week, Clinton defended the technology development programmes that have been close to the heart of his administration, and also praised what he called the "persistence and focused intellectual energy" of American scientists.

Before awarding 16 National Medals of Science and Technology, the president pointed out that the budget plans of the Republican-dominated Congress "will cut vital research and development by a third". He added: "We may have a balanced budget to show for it tomorrow, but a decade or a generation from now our nation will be much poorer for doing that."

Much of what Clinton said was familiar — for example, that it is important to strengthen, rather than weaken, US investment in science, technology and research. "We must sustain our universities — a critical national resource and still the envy of the entire world."

But he gave his remarks a new significance by seeking to tie investment in science and technology to the race issue, two days after the Million Man March of black Americans in Washington, which has highlighted the persistence of racial division as America's foremost political problem.

Drastic cuts in science and technology spending need to be resisted, he said, because "constant, churning innovation" is the key to economic growth and national strength in the twenty-first century. "If we're going to make real the promise of the American dream to all Americans — which would plainly do a lot to help us deal with the kind of racial difficulties that we began so bravely as a nation to come to grips with this week — we have to go further in this area."

Clinton's much-heralded science speech contained few clues as to how far he is prepared to go in resisting proposed Republican cuts for the 1996 financial year just begun, which fall most heavily on technology programmes such as the Department of Commerce's Advanced Technology Pro-

gramme and on environmental research.

Indeed, it is not even known when the President and the Congress will finally agree on a 1996 budget; they were supposed to do so by 30 September, and the rescheduled deadline of 13 November is now widely expected to slip. The cuts of one-third mentioned by Clinton are based on estimates by his Office of Science and Technology Policy of the impact of Repub-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Helping hands: Vice President Al Gore (left) and President Clinton congratulate medal winner Louis Nirenberg

lican plans to balance the budget by the year 2002.

The eight scientists honoured by the president included the chemist Isabella Karl, head of the X-ray diffraction section of the Naval Research Laboratory, who was selected for her development of techniques, now used worldwide, for the X-ray analysis of crystal and molecular structures. Louis Nirenberg, former director of the Courant Institute of Mathematical Sciences at New York University, received a National Medal for his overall contribution to pure and applied mathematics.

The other winners were Thomas Cech of the University of Colorado at Boulder, for work with ribonucleic acid (RNA) enzymes; Hans Dehmelt of the University of Washington, Seattle, who built electromagnetic traps to capture and study subatomic particles; Peter Goldreich of Caltech for planetary science and theoretical astrophysics; Hermann Haus of the Massachusetts Institute of Technology (MIT) for quantum optics; Alexander Rich of MIT for structural biology; and Roger Shepard of Stanford University in California for work on mental imagery.

Clinton and his vice president, Al Gore, also gave out eight Technology Medals. These included one to the company 3M, whose development of the Post-It note was singled out for special praise by the leader of the world's leading industrial economy.

Colin Macilwain

Brussels bids for Framework 'top-up' funds

Munich. Edith Cresson, the European Union (EU)'s research commissioner, last week won the support of her 19 fellow commissioners for a bid to allocate ECU700 million (US\$925 million) in potential 'top-up' funds to the work of five task forces set up to promote the development of new technologies in Europe under the EU Framework programme.

The move could end the hopes of Europe's national research councils that some of this money would be used to improve the coordination of their basic research with the goals identified by the task forces, and to provide large awards to a selected group of leading research teams, comparable to Germany's Leibnitz prize.

But while disagreement remains over how the money should be used, there is also concern about whether it will in fact materialize. The proposed top-up fund represents ECU600 million to be held in reserve for the fourth Framework programme, which runs from 1995 to 1999, and ECU100 million for the research programmes of the atomic energy agency Euratom.

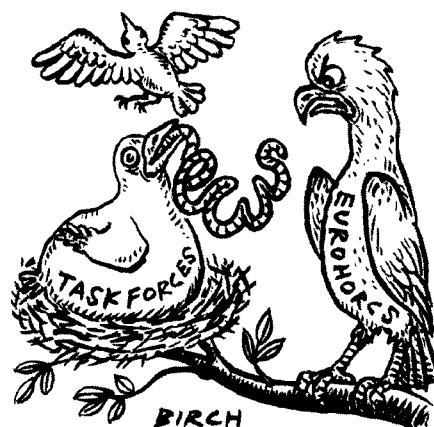
The precise allocation of the money, initially left undefined on the grounds that it would be used to fine-tune spending on both programmes, has to be approved by the commission, the European parliament and the European Council next year. Some finance ministers are already arguing for a sharp reduction.

Cresson, in contrast, is already firm on how she would like the full money to be spent. Together with Neil Kinnock and Martin Bangemann, commissioners for transport and industry respectively, she has set up task forces to investigate how better

coordination of EU research could promote the development of new technologies in areas such as the 'clean' car and vaccines against viral diseases (see *Nature* 374, 206; 1995).

The development of safe water supplies was recently added to the portfolio for the benefit of Mediterranean countries, and it is planned to feed the results of the task forces' work into the fifth Framework programme, which begins in 1999.

Some of the task forces have already made significant progress, and coordination plans are being put into practice. Earlier this month, for example, Cresson announced a three-month delay to the second call of pro-



posals for the technology-orientated ECU1,200 million Brite-EuRam research programme, to line up with the December call for the industry commission's major ESPRIT programme.

The delay will give managers of the Brite-EuRam programme time to realign its scope with the ideas of the task forces concerned with transport. A spokesman for the programme says that focusing its aims will have a second benefit; its original broad remit meant that the programme was heavily oversubscribed in the first call for proposals, and the acceptance rate was only 18 per cent.

The commission argues that EU member states should pay the top-up fund in full because of the need to coordinate research in specific fields to ensure the general competitiveness of European industry. Priorities include the next generation of cars and aircraft, multimedia education technologies, integrated transport systems and environmental technologies, including the protection of water supplies and a Euratom programme on the security of eastern European nuclear reactors.

The broad proposal will be considered next week by the research ministers of EU states. But it may have to be adjusted before being put to the full European Council at the end of the year if the finance ministers, meeting next month, refuse to approve the full sum.

Meanwhile, the heads of Europe's national research councils — known collectively as Eurohorcs — meeting in Venice earlier this month, decided to seek the support of the European Science and Technology Assembly (ESTA), the commission's 100-strong advisory panel, for its own ideas of how the top-up money should be spent.

The Eurohorcs are worried that the task forces' industrial aims may result in basic research being ignored, and would therefore be prepared, with the commission, to finance basic research projects relevant to the task forces. The exact nature of the coordination would depend on how much of the funding for the proposed projects would come from the commission and how much from the research councils themselves.

At the suggestion of Wolfgang Fruhwald, president of Germany's Deutsche Forschungsgemeinschaft, they would also like ECU20 million set aside for ten awards to leading research teams, a scheme already successfully applied in Germany.

The research council heads argue that both schemes would improve the scientific credibility of the research commission's work. But at its meeting last week, ESTA declined to decide whether to support the two proposals because of uncertainty over the amount of top-up funds, according to the assembly's chairman, Jan Boorgman.

A spokesman for Cresson's cabinet says that she was prepared to consider any formal proposal for the use of the top-up funds, but doubted whether she would agree to split the money into relatively small sums.

Alison Abbott

EMBO seeks stronger voice on policy

Heidelberg. The European Molecular Biology Organization (EMBO) wants to raise its profile in Europe by reorganizing its 850 individual members into a type of academy of sciences able to contribute a collective voice to debates on European science policy.

Earlier this month, EMBO's council agreed that it needs to make more use of its members in helping to promote molecular biology in Europe, primarily through its increasingly popular short- and long-term fellowships, as well as specialist courses and workshops.

At present, the main role of EMBO members, who are elected on the basis of academic achievement, has been to interview candidates for fellowships. The council would now like to hold annual meetings of its members to discuss science policy issues affecting molecular biology.

The EMBO council also agreed to pay for EMBO members wishing to help

increase the interest of east European countries in taking part in its activities. The number of applications for EMBO's special programme of six-month fellowships for east Europeans has been falling steadily.

According to Gannon, this may be because inadequate facilities means that returning scientists are unable to put into practice what they have learnt in Western Europe. EMBO is to invite its members to visit molecular biology laboratories in such countries for a week, to try to improve the match between the specific needs of the laboratories and appropriate projects in Western Europe, and to help the laboratories to apply successfully for foreign grants.

EMBO will also finance plenary lectures at national and international symposia, particularly in east European countries, in order to stimulate support for the European Molecular Biology Laboratory (EMBL) in Heidelberg.

A.A.

ESA avoids split over space station plans ...

Toulouse. France and Germany last week pulled Europe's joint space programmes back from the brink of collapse in a flurry of late-night negotiations that led to agreement on Europe's participation in the international space station, and the continued development of the new Ariane-5 launcher.

The negotiations marked the nail-biting climax to a meeting of the ministers of the member states of the European Space Agency (ESA) in Toulouse, France. "Europe is back in orbit," declared a beaming Ivan Illieff, Belgium's science minister and the chairman of the three-day meeting, as he announced that consensus had been reached on all the crucial decisions that it had faced.

At Illieff's side, looking grey-faced from lack of sleep, Jean-Marie Luton, the director general of ESA, smiled nervously. "The money didn't exist, but it came," he sighed with relief, referring to the end of months of deadlock among the member states over funding of the agency's proposals.

Failure to break this deadlock would probably have led to the disintegration of ESA, claimed François Fillon, the French minister responsible for space, with Germany withdrawing in order to participate on its own alongside the station's other international partners, the United States, Russia, Canada and Japan. "It could have been handkerchief time tonight," he remarked.

But agreement has not been without its price. Both France and Germany have significantly increased their commitments to the costs of the programmes approved at the meeting. At the same time, under the agreements reached at the meeting, ESA itself will have to reduce its annual running costs of ECU170 million (US\$225 million) by 12 per cent over the next five years, and reduce its 3,500 staff by 500.

The meeting agreed to spend ECU2.8 billion on Europe's contribution to the space station between 1996 and 2004. This will include development of the Columbus Orbital Facility (COF), to be built by Germany and Italy, an Automated Transfer Vehicle, built by France, to ferry freight and fuel to the station.

The meeting also agreed that Europe's contribution to the running costs of the station should be made in kind, and not exceed the equivalent of one Ariane-5 flight to the station every 20 months, relieving fears among the member states that the running costs of the station might erode ESA's budget. Although this proposal will need to be approved by the other international partners, Daniel Goldin, administrator of the US National Aeronautics and Space Administration, told Luton this month that it would be acceptable.

Agreement on the financing of Europe's space station contribution was primarily the

result of France and Germany — already ESA's largest paymasters — agreeing to support each other's pet projects.

Germany's main interest is in the space station, and France, which had previously agreed to withdraw a proposal to substitute a French built crew rescue vehicle (CRV) for the planned COF space laboratory (see *Nature* 377, 191; 1995), agreed to increase its funding for the German-led station programme from 16 to 27 per cent. Germany will pay 41 per cent of the costs of the programme.

At the same time, recognizing France's goal of maintaining the autonomous access of Europe to space through the Ariane launcher programme, Germany has agreed to increase its contribution to the French-led programme to further develop the Ariane-5 launcher from 7 to 17.7 per cent. The meeting agreed to spend ECU1,721.4 million on this programme between 1996 and 2003, with France meeting almost half the costs. In a further gesture towards France, the meeting also agreed that ESA will spend ECU50 million on further technical studies of its CRV project.

Despite these concessions, negotiations almost broke down when Italy repeated its previous refusal to pay more than ECU100 million of the ECU300 million it had originally promised to ESA's planned contribution to the station over the period 1996–2000. Eventually, however, Italy agreed to borrow ECU150 million to make

up its contribution, after France and Germany had promised that Italian industry would be awarded an extra ECU50 million in contracts, and ESA agreed to make ECU50 million of economies in the programme.

As a result, Italy will now contribute 19 per cent of the costs of ESA's contribution to the station, and 6 per cent of the costs of the Ariane development programme. The United Kingdom, which has never participated in the Ariane programme, also indicated that it would contribute 2 per cent to the launcher programme.

There will, however, be a heavy price to pay by France and Germany — as well as ESA itself — for the space station decision. The French space agency, CNES, for example, will have to cut its budget by 5 per cent to meet France's obligations.

Jacques Chirac, the French president, applauded the agreements reached. But Lionel Jospin, the leader of the opposition socialist party, criticized the level of France's increased contribution to the station programme, warning that it could damage national programmes.

But the major political outcome of the meeting seems to be the firm re-establishment of the

Franco-German axis as the central pillar of Europe's space programmes. Papering over the tensions of the past few months, when Germany and France were backing their rival versions of Europe's contributions to the space station, Fillon said that a critical factor had been a meeting of minds between the two countries, a sentiment echoed by Jürgen Rüttgers, his German counterpart.

Declan Butler



Going up: Ariane V will help to pay space station bill.

... but holds back on space science

Toulouse. The science programmes of the European Space Agency (ESA), considered by many as 'the jewel in ESA's crown', have fallen victim to twin pressures of a general squeeze on the space budgets of some of its member states due to their commitments to the space station (see above) and pressure from some states to shift the balance of spending back to domestic and bilateral programmes.

In the run-up to last week's ESA space ministers meeting in Toulouse, ESA had been insisting that the budget for the various programmes in its flagship Horizon 2000 science programme should be increased at the rate of inflation over the next five years. But

the United Kingdom had opposed this increase, arguing that ESA could cut 25 per cent of the costs of its science missions over this time period, while Germany proposed that the budget remain level.

Under a French-brokered deal agreed at the meeting, the annual budget of the science programme will be kept constant at ECU347 million (US\$457 million) from 1996 to 1998 — roughly in line with German demands — and will be indexed only for any inflation over 3 per cent. The meeting agreed to reconsider in 1998 whether a constant budget should be maintained up to the end of the decade.

Roger Bonnet, head of ESA's science ►

programmes, has so far declined to comment on the reduction in the science budget until its full impact has been assessed, although he is reported to be "extremely disappointed".

David Southwood, professor of physics at Imperial College, London, and chairman of ESA's science programme committee, describes the prospect of a 9 per cent cut over three years as a "squeeze" that can be managed. But he goes on to say that he would be worried if the level of cuts were extended beyond then, arguing that this would lead to a general attrition of the programme.

The United Kingdom, which will be in a position to veto any further increase after 1988, at present feels confident that the level funding will be extended to the full five years. Ian Corbett, director of science at the UK Particle Physics and Astronomy Research Council (PPARC), says this will result in a reduction of 15 per cent in ESA's space science budget — equivalent to a saving of £15 million for the United Kingdom that is now available for its domestic space programme — and is therefore "not much less" than the 25 per cent reduction which Britain had been seeking.

Indeed, Britain's delegation to the ministerial meeting described the outcome as a "success". According to Corbett, as well as making money available for PPARC to spend on instruments and exploiting the results of missions, the agreement will improve PPARC's ability to plan in the long term, as ESA will no longer have to find extra money each time the inflation rate changes. "The arguing is now over," says Corbett. "We and ESA have a stable basis on which to move forward."

Jean-Marie Luton, the director general of ESA, claimed during last week's meeting that the budget cuts were unlikely to have a significant impact on planned space missions. But others are less optimistic. One agency official points out that all of ESA's programmes are under financial pressure, and that the science programme could still become a target for further cuts.

Such cuts may not be made directly, but rather may emerge as the result of administrative reforms at the agency. The science programme, for example, has until now benefited from privileged charging arrangements for overheads, and these risk being trimmed back, according to Southwood.

According to Southwood, the overall challenge to the agency, and indeed to Europe's space industry in general, is to improve the efficiency with which it carries out projects. He refers in particular to proposed reforms of ESA's industrial policies, which will be reviewed at an informal ministerial meeting next year. "Industry will need to do more for less," he says. But he warns that delays — or even cuts — to the science programme now seem inevitable.

D.B.

Health bodies urge backing for early warning procedures

Washington. The World Health Organization (WHO) and the US Centers for Disease, Control and Prevention (CDC) have appealed to their respective funding bodies for more money to pay for joint and separate initiatives to identify, control and eventually eradicate emerging infectious diseases.

The WHO has asked its 190 member states for \$5.5 million to help foot the \$7 million costs for a worldwide early warning system to pre-empt infectious diseases, citing examples such as last year's spread of bubonic plague in India which killed 54 and left losses worth an estimated \$1.5 billion, and the Ebola virus outbreak that killed 300 in Zaire. The proposals include the setting up of a network of rapid response units to place expert health teams at any world disease zone within 24 hours of an outbreak.

Meanwhile the CDC is urging the US Senate to recommend an increase in its budget to match the \$125 million annual operat-

re-emerging disease through innovative field epidemiology and public health laboratory programmes," Heymann said last week.

The CDC project, which shares many of the WHO programme's objectives — although restricted to the United States — has so far received just \$7.7 million this year. The House of Representative has recommended an increase of \$12.8 million to the CDC's overall budget. But so far, the Senate has been less generous.

"History tells us that infectious diseases will remain important, evolving complex public health issues," David Satcher, the director of CDC, told a hearing of the Senate Labor and Human Resources committee last week. The main problem, he says, lies at the state and local levels, where funding cuts in health departments have left them "poorly prepared" to identify and respond to infectious agents. But Satcher added that federal agencies, led by the CDC, are well placed to deal with emerging infectious diseases both in the US and abroad.

His comments contrast with those in a 1992 report by the Institute of Medicine which criticized the nation's ability to identify and react to harmful organisms. This was seen as a clarion call for renewed effort against emerging infections.

The fact that the CDC and WHO projects share similar goals and the same lexicon — "to strengthen surveillance, strengthen prevention and control, strengthen infrastructures, foster applied research" — is no coincidence. One quarter of WHO's budget comes from the US government, and the US contribution is currently threatened with cuts of 35 per cent proposed by a Senate appropriations panel. WHO officials would dearly like to see CDC succeed in its bid for more funds. A vote for the CDC project would, by implication, be a vote for WHO.

Much of WHO's promotional literature addresses both its international as well as its US policymaker audience. In a briefing paper, Hiroshi Nakajima, WHO's director general, says the agency's strategy for tackling emerging infectious diseases "is consistent with the plan developed by the Centers for Disease Control and Prevention".

Meanwhile, a group of 17 US agencies led by the CDC, known as the Committee on International Science, Engineering and Technology, has been discussing how to upgrade the global response to threats of infection. In a report published in July, it recommends establishing a global surveillance network linking up to the CDC. A task force from this group will meet in November to produce another plan outlining how its recommendations should be implemented.

Adrianne Appel & Ehsan Masood

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

First signs: rats in Surat provided indicators of the possible arrival of bubonic plague.

ing costs once the conclusions of its report, *Addressing Emerging Infectious Disease Threats: A Prevention Strategy for the United States*, are fully implemented, planned for the year 2002.

The WHO project will be run by its newly created Division of Emerging, Viral and Bacterial Diseases Surveillance and Control (EMC). It will be headed by David Heymann, the physician credited with stopping the spread of Ebola to 2 million people in the Zairean capital of Kinshasha.

"EMC will work to strengthen country surveillance and disease control in order that countries that develop the early warning systems necessary to detect an emerging or

Galileo 'will still meet most scientific goals'

Washington. Mission managers handling the National Aeronautics and Space Administration's (NASA's) Galileo spacecraft are optimistic that their two-year tour of the Jupiter system can still meet most of its planned scientific goals — even if the spacecraft's faulty onboard tape recorder is declared a total loss.

Instructions transmitted to Galileo last weekend showed that the recorder, which began experiencing problems on 11 October (see *Nature*, 377, 563; 1995), is still able to move and read tape normally. But whether it can be trusted with critical images and data during the main mission — which begins on 7 December, when a probe is set to plunge into Jupiter's clouds just as the main spacecraft enters orbit around the planet — could be in doubt.

"We're assuming at the moment that we may well have to live without the tape recorder," says Torrence Johnson, a project scientist at NASA's Jet Propulsion Laboratory (JPL). Galileo managers are already making contingency plans for a 'real-time' mission, with science data either being relayed to Earth immediately, or stored in the spacecraft's computer memory, rather than on tape. According to this plan, additional computer memory would be made available for storing science data by using sectors that had previously been used to operate the recorder.

Such a strategy will reduce the number of bits of data that can be relayed back to Earth by as much as three-quarters. But Michael Belton of the Kitt Peak National Observatory, who heads Galileo's imaging team, claims that relatively little will be lost. "The science is not linearly proportional to the number of pixels," he says.

Indeed, Johnson estimates that even without the use of a tape recorder, Galileo will still be able to meet more than half its scientific objectives. The loss of the spacecraft's high data-rate antenna in 1991 — the cause of heavier reliance on the tape recorder — had already reduced the mission to 70 per cent of its original output. "The overall mission philosophy is nowhere near as much changed by this as it was by the

NIH radiation incident. A story last week reported that the National Institutes of Health had calculated that 27 researchers involved in an incident at the National Cancer Institute this summer had absorbed doses of radiation of between 300 and 600 microcuries (see *Nature*, 377, 568; 1995). In fact, NIH measurements showed that, apart from one researcher who had been exposed within this range, none of the others received doses of more than 25 microcuries. The misunderstanding arose from ambiguities in a draft statement inadvertently released by NIH.

high-gain antenna failure," he says.

In the revised plan, the mission's top priority, data from the atmospheric probe, would still be retained, as this was already to be stored in computer memory. Data on Jupiter's electromagnetic environment, most of which will be relayed back to Earth in real-time, would be only slightly affected.

Lost in the new plan, however, would be some of the photo opportunities generated by the spacecraft's arrival near Jupiter, including high-resolution pictures of the volcanic moon Io and of the southern hemi-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Getting closer: an impression of Galileo's encounter

sphere of Europa. Unless the tape recorder can be made operational — or additional space can be freed in computer memory — the probe data will take up all the available storage room.

Observing sequences for the rapid events surrounding arrival at Jupiter will remain unchanged from the original plan, as it would be too risky to make major changes with only two months left. Any revisions to the tour would begin next July with the satellite fly-by of the moon Ganymede.

During these satellite encounters, says Belton, the imaging team will lower the

number of bits of data by reducing the frame size for each image, so that some areas will not be photographed at high resolution. The strategy will still allow areas of scientific priority to be photographed at resolutions of hundreds, and in some cases tens, of metres, about 100 times better than those produced by the Voyager fly-bys of the 1980s. "What [the recorder problem] translates into is a reduction in [the area covered], but the resolution is maintained," says Belton.

He and other mission scientists say they are in better spirits than the dark hours immediately after the tape recorder problem first surfaced. "You should have heard the cheer that went up around our teleconference after we started getting the report [of how much data could still be retrieved]," says Belton. "We thought we were dead in the water."

The Galileo team is no stranger to adversity. The spacecraft, which was built more than a decade ago, was expected to complete its Jupiter mission in 1990. But the explosion of space shuttle Challenger four months before its planned launch in May 1986 threw that schedule out the win-

dow. Other shuttle-related problems meant that Galileo had to take a slower, more circuitous route to Jupiter after its launch in 1989, and will now arrive at its destination seven years late.

While most of Galileo's problems are not of its own making, the apparent failure of the mission's single tape recorder — installed despite calls for a back-up — may provide a cautionary tale for today's budget space missions. If hardware failures continue, the best that project managers can do is to cross their fingers and keep devising clever alternatives.

Tony Reichhardt

SKB backs genome ethics programme

San Francisco. The pharmaceutical company SmithKline Beecham, which entered into a \$120-million agreement in 1993 with the gene sequencing company Human Genome Sciences, has donated \$950,000 to Stanford University in California to fund a research programme on Genomics, Ethics and Society over three years.

The programme, which also hopes to draw funding from other sources, plans to take a multidisciplinary approach to the issues raised by genetic testing, gene therapy and other technologies resulting from the advance of human genetics.

Initially, a working group made up of representatives from the humanities and the sciences, as well as pharmaceutical companies, patient groups and health insurance providers, will review policy in the area of

genetic screening for susceptibility to breast cancer. The working group will prepare a 'white paper' (policy document) leading to an international conference to be held in the autumn of 1996 at Stanford.

Barbara Koenig, the programme's director, says it will take a critical stance on the development of genomics, building on a foundation of scientific expertise. There are no reporting requirements to SmithKline Beecham — which is familiar with controversy surrounding genomics as a result of its funding of HGS — or any restrictions on publication.

But to avoid any apparent conflict of interest, the programme will not consider questions raised by patents or other issues directly related to SmithKline Beecham's product areas, says Koenig. **Sally Lehrman**

Expert panel to judge on deep sea disposal

London. Tim Eggar, Britain's minister for industry and energy, has set up an international panel of scientists and engineers to evaluate the technical issues raised by the disposal of the Brent Spar oil storage platform, and to appraise new proposals put forward by the oil company Shell.

The panel will be led by the government-funded Natural Environment Research Council (NERC), which favours deep-water disposal of Brent Spar, currently moored off the Norwegian coast following a decision by Shell, its owners, to abandon previous plans to dump the platform in the North Atlantic (see *Nature* 376, 378; 1995).

Both the membership and terms of reference of the committee have yet to be finalized. But Eggar, who opposes the on-shore dismantling of the Brent Spar, said last week that deep-sea disposal would remain the "environmental benchmark" against which other options would be judged.

The announcement was made after an independent audit of the Spar's contents challenged a previous estimate by the environmental group Greenpeace — which was primarily responsible for Shell's earlier decision to abandon deep-sea disposal — of the amount of oil it contained.

The audit, performed by the Norwegian certification agency Det Norske Veritas (DNV), calculated the oil content at between 75 and 100 tonnes, relatively close to Shell's estimate of 53 tonnes. In addition, despite Greenpeace's claims, it found no evidence of toxic materials, and negligible quantities of radioactive waste.

But the impact of the report was immediately undermined by evidence of an internal split within Shell over the company's next step, when two senior officials made apparently contradictory statements about the

Brent Spar's eventual fate.

At a press conference in London following the audit results, John Wybrow, Shell's director of corporate affairs, said twice that deep-sea disposal of the platform had been ruled out. But this was subsequently described as "premature" by Eric Faulds, manager of the Brent Spar audit project.

Wybrow claimed the audit results "put to rest" Greenpeace's "grossly overestimated" calculations of the contents of the platform,

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Where next? Shell has promised to listen to public opinion over the ultimate fate of the Brent Spar

and stressed the company would now concentrate on "finding a new solution", which "manifestly rules out deep-water disposal". The company has so far received 250 proposals after inviting suggestions for how disposal should be achieved.

But Wybrow also acknowledged that Shell had made a mistake in ignoring public concern about deep-water disposal, saying that any future decision to dispose of a disused offshore oil installation should not be based solely on scientific, economic and environmental considerations. "Political and public considerations will take greater

priority [than at present]," he said.

Nevertheless, Shell continues to insist that deep-water disposal remains a viable option. A spokeswoman for the Department of Trade and Industry said later that Chris Fay, the chief executive of Shell UK, has confirmed to Eggar that "deep-water disposal remains an option for Brent Spar", particularly in the light of the audit results.

A campaign by Greenpeace — including coordinating a Europe-wide boycott of Shell petrol-filling stations — forced Shell to abandon plans to sink the Spar on 20 June. Greenpeace claimed the sinking would amount to an environmental catastrophe.

Wybrow's proposal — confirmed by a Shell spokesman — to allow the public to influence a decision on Brent Spar, is likely to be controversial, and may not be supported by the government.

A spokeswoman for the DTI says the department would not speculate on the relative importance given to public acceptability until a proposal from Shell is on the table. But Greenpeace welcomed the development. Any decision to dispose of Brent Spar must take account of public opinion, says its spokesman, Adam Woolf. "Science plays a vital role," says Woolf. "But it cannot be the sole arbiter of truth and wisdom, because it is often value-driven, too."

The argument about the fate of the Spar, has overshadowed what is arguably an equally important issue: Greenpeace's ability to influence public opinion using information that, according to Ole-Andreas Hafnor, senior vice-president of DNV, "contained major errors of misinterpretation".

Although the DNV report points out inaccuracies in both Shell and Greenpeace data, Greenpeace's errors are more substantial. For example, the auditors found no evidence to support a sworn affidavit from a former Shell employee who claimed three barrels of toxic waste had been hidden on Brent Spar. The amount of "naturally occurring" radioactivity was too small to be classified as radioactive waste, Hafnor added.

But Woolf says that while Greenpeace acknowledged errors in estimates for oil on the Spar, the campaign was never about its contents. It was about setting "acceptable standards of behaviour" and "not treating the sea as a global dustbin".

Meanwhile, Robert May, the government's chief scientific adviser, said at a conference in Oxford last Friday that the matter underlined the need for a better early warning system within government to identify in advance issues such as those raised by the public controversy over the Brent Spar disposal.

Ehsan Masood

Astronomers seek a 'defence' agency

Boston. An international meeting of astronomers has agreed to promote the setting up of a network of telescopes to search for and monitor 'near-Earth objects' (NEOs) — asteroids or comets that may pose a threat to life on this planet — in line with a proposal made in 1992 by a panel set up by the US National Aeronautics and Space Administration (NASA).

The meeting, which was sponsored by the International Astronomical Union and held on the Italian island of Vulcano last month, proposed that this should be done through an international agency, to be known as the 'Spaceguard Foundation'.

"No existing agency currently supports this effort," says Eugene Shoemaker of the Lowell Observatory. Even the International Astronomical Union's NEO working group, headed by the Italian astronomer Andrea

Carusi, who also chaired the Vulcano proceedings, has only "a finite lifetime and essentially no budget". The foundation hopes to secure financial backing from government agencies, corporations and individuals, and then distribute funds to researchers. But Carusi says that raising money is not the main goal. "We want worldwide recognition that this is important work that needs to be done. With that recognition, the money should come."

"The key thing is to broaden the effort and make sure it's not just a US operation," adds David Morrison, the astronomer from the NASA's Ames Laboratory who led the original Spaceguard panel. "The foundation will focus on countries other than the United States, which already has an observation programme — albeit one that some people consider inadequate." **Steve Nadis**

Assessing track records

SIR — Your correspondents Giovanni Motta and Neil B. Metcalfe (*Nature* 376, 720; 1995) criticize journal impact factors (mean citations per paper) as a means of evaluating scientific track records. Although impact factors may on their own provide a flawed assessment of an individual's performance, more sophisticated analysis can provide useful and objective information.

At the Unit for Policy Research in Science and Medicine (PRISM) at the Wellcome Trust, we have been undertaking experimental analyses to ascertain whether it is possible to improve on the simple use of impact factors when assessing publication performance. Three considerations have driven us: (1) the fact that performance is not distributed normally (neither in citations, nor other areas of human endeavour); (2) the search for appropriate benchmarks against which individuals can be compared; and (3) a belief that multiple indicators of performance will be more reliable than single measures.

An essential feature of our analysis is the definition of a cohort of comparison papers. In particular, we generate sub-field-specific citation benchmarks that are appropriate for each applicant. We cannot rely on a subset of specialist journals, as many applicants work in a very narrow area and often publish their better papers in general journals. We therefore retrieve relevant papers by means of a complex title keyword filter, determined by consultation with scientists knowledgeable about the field. The comparison cohort should ideally comprise several hundred papers.

We then retrieve articles, notes and reviews by an applicant from the *Science Citation Index* and characterize each publication by the number of pages (P), the number of authors (A) and a journal weighting on a scale of 1 to 4 (the best) (W). A consideration when a journal is being weighted is that the typical citation distribution is distinctly nongaussian, making the mean (as used for impact factors) an unreliable measure of central tendency (Fig. 1). We therefore employ categorical analysis, where journal weighting depends on how its citations relate to those of the cohort of comparison papers. We then calculate fractional outputs for each paper in the form of P/A , W/A and WP/A .

We also determine the C_{0-4} values (citations received over a five-year period) for each research paper on which the applicant is an author. These values are then compared with the distribution of C_{0-4} values for the comparison cohort in order to determine in which quartile each article lies. An article in the top

quartile is given a CQ value of 4, and an article in the bottom quartile is given a CQ value of 1. Papers in the top 10 per cent and top 5 per cent are given additional annotations.

The applicant's list of publications can also be compared with the track records of other leading scientists in the same scientific area. A useful indicator here is WP/G , where G is the number of authors

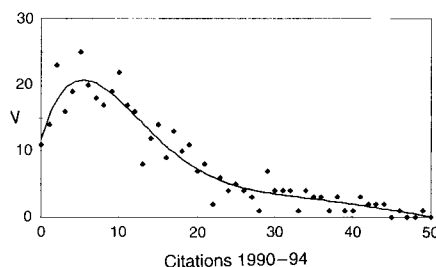


FIG. 1 The distribution of C_{0-4} values (citations received by each paper over a five-year period) for the 424 papers published in 1990 in the journal *Neuroscience*: the mean is 16.4 but the median is 11 and the maximum is 340.

in the group with whom the applicant has copublished. This measure shows scientists who either have large groups of their own or collaborate extensively with others, and may thus obtain credit for administrative rather than scientific brilliance.

We believe that our group of publication indicators can provide grants committees with useful input that they can then interpret in the light of their specialist knowledge in order to form an overall judgement on the scientific track record of applicants. All our analyses are based on the premise that publication lists are reasonable measures of track records, which is not necessarily true. Nevertheless, the publication lists of applicants continue to be referred to by grants committees when they are making awards. If inspection of publication records is to continue as part, but only a part, of the peer review process, the analysis must be carried out with the same degree of intellectual rigour that would be expected in the conduct of good scientific research. The alternative is to risk undue influence from subjective impressions, hearsay, anecdote and simple prejudice.

Grant Lewison

Joe Anderson

PRISM,

The Wellcome Trust,
210 Euston Road,
London NW1 2BE, UK

Julian Jack

University Laboratory of Physiology,
University of Oxford,
Parks Road,
Oxford OX1 3PT, UK

Mind mechanisms

SIR — Neuroscientists now study consciousness and will probably soon investigate the molecular biology of the mind. In parallel some argue that the mind houses mechanisms that enable people to develop multiple personalities, a sign of psychopathology that has increased dramatically recently¹. There is no scientific foundation for this development¹ and one wonders how it is possible now when molecular biology permeates clinical medicine.

It is unlikely that science can benefit from non-scientific developments. However, the reverse seems true. Thus, a representative of the non-scientific approach feels: "... on common ground with molecular biologists. Working with the multiple personality disorder patient, one has a sense of ... looking directly at its (the mind's) basic structure..."¹.

The stimulatory effect of growing scientific knowledge on non-scientific approaches to the questions of life is well known but has not been explained. Perhaps it reflects our ignorance about how one mind interacts with another, that is our limited understanding of interpersonal relations. Knowledge of these might solve some mental health problems, because the relationship between patient and mental health professional is distorted. In a remarkable study published 22 years ago, mental health professionals could not distinguish the mentally ill from those with no mental illness. By contrast, the mentally ill could². Although this study was not followed up, distinguishing patient from pseudopatient may remain a problem because 75 per cent of the symptoms in the commonly used diagnostic manual of mental disorders are examples of normal behaviour³.

Understanding interpersonal relations is the 'big question' now facing biologists. Neuroscience and molecular biology may help in answering the question, but new methods will be needed, suggests a molecular biologist⁴.

Per Södersten

Karolinska Institute,
Department of Clinical
Neuroscience,
Novum, S-141 57 Huddinge,
Sweden

1. Offshe, R. & Watters, E. *Making Monsters* (Schribners, New York, 1994).
2. Rosenhan, D.L. *Science* 179, 250-258 (1973).
3. Pendergrast, M. *Victims of Memory* (Upper Access, Inc., Hinesburg, 1995).
4. Baltimore, D. *Genes & Dev.* 9, 1303-1307 (1995).

Correction: Eradicating nuclear weapons

The last sentence of the letter by Jan H. J. Oelering (*Nature* 377, 10; 1995) should have read: "So rather than a few months after Hiroshima, publication [of the Smyth report] took place a few days after Hiroshima". □

Duesberg and AIDS

SIR — In the wake of strong evidence for the causal link between HIV and AIDS, a recent leading article suggested to those who oppose this theory that “there may come a point at which the dissenters forfeit the right to make claims on other people’s time and trouble by the poverty of their arguments and the exasperation they have caused”¹.

This argument strikes me as rather spurious. If Duesberg has a logical argument that suggests the mainstream view is wrong, it should be heard. If the argument is poor, it will cost little time and trouble, and cause little exasperation. If it is so brilliant as to be irrefutable, it will also cost little time and trouble, and cause little exasperation, being simply accepted. It is only in the grey area in between that scientists will be forced to spend time and trouble and are liable to become exasperated. This grey area, however, is where most scientists spend most of their time. If one does not like spending time and trouble on another’s ideas, and becoming exasperated by them, then one is in the wrong job.

Since the publication of Duesberg’s seminal paper suggesting that retroviruses are too weak to cause illness², scientists have indeed spent time and trouble in dealing with him. Sadly, in their exasperation, they have used bans, censorship and personal attacks to keep Duesberg silent. He has been ignored, had his funding cut and been denounced in *Nature* as not belonging to “the grown-up world”³.

Duesberg and his colleagues (the ‘dissenters’) have faced peculiar multilayered review boards when submitting their papers to scientific journals — producing far greater delays and higher rejection rates than normal — while articles that refuted Duesberg’s claims have been ‘fast-tracked’, by-passing the normal peer review process. The dissenters have also been excluded from mainstream scientific meetings, forced to meet separately like a clandestine cult. Such unorthodox, and unacceptable, manoeuvres beg the question why.

Duesberg’s ideas have received such widespread attention largely because of the failings of the mainstream AIDS establishment. The moral claptrap that has surrounded AIDS since its inception established AIDS as a major killer before one scrap of evidence was obtained. The expectation of AIDS as a widespread virulent killer, a view that most scientists held until very recently, meant that simplistic

ideas as to how HIV worked were liable to be favoured, simplistic ideas that Duesberg was able to attack. So Duesberg could snigger as AIDS failed to penetrate the heterosexual community, smirk as scientists tried to explain the ever-increasing latency of HIV infection to AIDS, and point to his own theory as being closer to the pattern of AIDS development than anything offered by the mainstream.

It was the belief that AIDS was going to explode into the heterosexual community that took up so much time and trouble, not Duesberg. As this misconceived notion fades away, a new science of HIV can emerge, one that will not include Peter Duesberg, but which may nonetheless be troubled by him as a penance for past miscalculation.

Stuart W. G. Derbyshire

*Rheumatic Diseases Centre,
Clinical Sciences Building,
Hope Hospital, Eccles Old Road,
Salford M6 8HD, UK*

1. *Nature* **377**, 1 (1995).
2. Duesberg, P. H. *Cancer Res.* **47**, 1199–1220 (1987).
3. Maddox, J. *Nature* **363**, 109 (1993).

Ill-fated congress

SIR — The Fifth International Congress of Genetics was held in Berlin in 1927, the Sixth in Ithaca, New York, and the ill-fated Seventh was originally scheduled for Moscow, not Nazi Germany as a recent News story claimed (*Nature* **377**, 7; 1995).

It was N. I. Vavilov who suggested that it be held in Moscow and he was active in arranging the meeting for 1937 when H. J. Muller arrived and worked with Vavilov to organize the conference. Unfortunately the young T. D. Lysenko was gaining influence, and V. Molotov helped Lysenko in adding Lysenkoists to the planning committee and by making “impossible” demands, such as requesting that there should be no human genetics as that was inherently racist and that there should be no German delegation. The delays and the attacks on genetics by the Lysenkoist camp led to the rescheduling of the congress (for 1939) in Edinburgh. Vavilov hoped as late as 1937 that the congress could be held in Moscow to serve as a bulwark against the growing Lysenkoist movement.

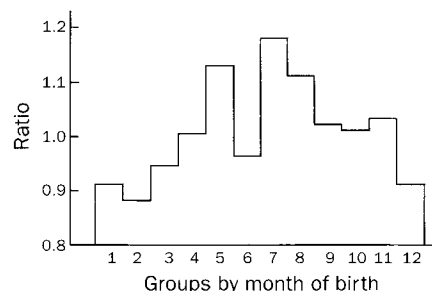
The seventh congress was indeed ill-fated. The war broke out as the congress ended and a ship (the *Athenia*) in which some delegates were returning to the United States was torpedoed and a few geneticists lost their lives (most were saved).

Elof Axel Carlson

*Department of Biochemistry
and Cell Biology,
State University of New York,
Stony Brook, New York 11794-3351, USA*

Born in summer?

SIR — Among the students who entered Porto faculty of medicine in the past three years, Azevedo *et al.* observed a significantly higher ratio of students born during the second trimester of the year and discussed possible effects of birth dates on early development of the brain (*Nature* **376**, 381; 1995). Although their result is interesting, the number of the students



examined (263) seems too small to establish the observation.

I investigated the birth dates of 2,525 graduates from the faculty of medicine at the University of Tokyo during the past 25 years and grouped them by month. These graduates were also among the most successful high-school students in terms of achievement in written examinations. In each group, the numbers of graduates observed were divided by those expected from births by month in Japan during 1947–71 when the graduates were born. The average ratios are shown in the figure. There is a significant difference in the ratios between groups born during the summer and winter, with the highest in July and the lowest in February ($P < 0.01$). Of the four trimesters of the year, it is not the second but the third trimester that shows the highest ratios, which cannot be explained by a relative age advantage because April is the start of the academic year in Japan.

My result does not completely agree with the observation of Azevedo *et al.*; nevertheless, it is also suggestive of the possible effects of birth dates on academic success. According to my result, birth during summer seems advantageous. Why summer? Although the reason is unclear, one possible explanation is that babies born in summer spend their first couple of months under the least restraint of heavy clothes and of illness and with the opportunity of staying outdoors, which would increase their physical movement and openness to external stimuli, both probably important for early development of the brain. This may also explain the disproportionately low ratio in the group of June, which is Japan’s rainy season.

Takanari Gotoda

*Molecular Medicine Group,
Hammersmith Hospital,
DuCane Road,
London W12 0NN, UK*

Correspondence

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only.

Ecocide in the Caspian Sea

Henri Dumont

As the Caspian Sea continues to rise, related environmental disasters are destroying not only local human communities but also valuable biological resources in the lake itself.

THE Caspian Sea is overflowing with problems. After decades of decreasing water level, the sea unexpectedly and rapidly began to rise in 1978 (see box over). Many local townships are now either flooded or on the brink of flooding. But that is just the start of the trouble.

Last April, I participated in a joint mission to the region, organized by the World Bank, the United Nations Development Programme and the United Nations Environmental Programme. We found that the rising level of the Caspian is a major problem for all the governments of the littoral states (Russia, Kazakhstan, Turkmenistan, Iran and Azerbaijan); it also became clear that, in the wake of the flooding, an unprecedented ecological catastrophe is developing, threatening to reduce the Caspian to a dead sea. The uniqueness of the situation lies in the multitude of simultaneous related dangers, most of them legacies of the former Soviet Union's obsession with eco-

the former Soviet Union as well as on poor ecological planning. We heard both arguments over and over again in all the littoral countries except Iran.

In Baku, oilfields are scattered from the hills surrounding the city to the sea's shallows in front of the town. Because of sloppy oil extraction and lack of maintenance, rigs and pumping stations are now littered with hundreds of oil lakes, which are swallowed by the rising water one by one. There are similar sights in all the five main oil-exploitation areas of the Caspian.

The Kura river is another source of pollution. Draining mining and industrial

the worst dangers in the Caspian basin, dwarfing the radioactivity in the Ural. Thankfully, unlike the Volga and the Kura, the Ural has not been extensively dammed, and is the only major river still used for sturgeon spawning migrations.

What of the ecology? Although not nearly as rich in life as Lake Baikal, the



Flooding near Cheleken.

Caspian is a remnant of the Tethys, which became freshened in the Pleistocene and evolved some 400 endemic species, such as the Caspian seal, still numbering about 500,000 individuals.

Among the 40 commercially exploited fish species (out of a total of around 120), the six species of sturgeon are unique. For about half a century, yearly

catches amounted to roughly 30×10^3 tonnes, and provided 90 per cent of the world's caviar. With the decline in catch numbers in the 1930s, quotas were enforced and catches stabilized at around 20×10^3 tonnes a year in the 1960s and 1970s. But dams on the Volga and the Kura reduced the number of sturgeon spawning grounds by about 85 per cent. Although fish lifts were constructed on the two lowest spawning grounds on the Volga, at Saratovskaya and Volzkaya, the number of spawning fish actually transported over the years is revealing: 60,000 in 1967, 29,700 in 1980, 3,400 in 1984 and 2,250 in 1987.

These lifts are now useless because of pollution of the river, which receives about 25 km^3 of waste water a year. Indeed, there were 17 recorded episodes of fish kills in 1987 and a major sturgeon kill (5,800 tonnes) in 1988; fish that survived had serious skeletal and muscle



Oil field on the coastal flat south of Baku.

nomic growth and the 'taming of nature'.

Take, for example, the town of Sumgait, northwest of Baku in Azerbaijan. Situated next to rich oilfields, the town underwent an industrial build-up in the 1950s and 1960s during the lake's recession. A complex of petrochemical, chemical and metallurgical plants, today all obsolete, decrepit and leaking waste into the lake, spans several kilometres of coastline. In 1991, Sumgait alone produced 335×10^3 tonnes of waste, mostly toxic (including dioxins). The local authorities, aware of the problem, blame it on the break-up of

areas in Georgia and Armenia, the river carries heavy metals such as copper and molybdenum and adds a cross-boundary dimension to the problem: part of the reason why Azerbaijan invests so little in environmental clean-up is due to its war with Armenia. For all practical purposes, the coast of Azerbaijan may already be considered dead.

In the desert of Kazakhstan, at Aktau, another potential disaster is emerging: radioactive contamination. The Gur'evskaya nuclear reactor uses Caspian water for cooling and desalination, but lies dangerously close to the shore, and discharges liquid waste into a natural depression roughly 8 metres above the Caspian's current level. Because this radioactive lake lies on karstic (limestone) soil, it is probably already leaking into the Caspian. There are also many solid nuclear waste sites in natural depressions, many in flood-prone zones. These may be some of

defects. To compensate for these losses, 17 sturgeon-production units have been built, seven on the lower Volga. Most release 3-gram fingerlings about 30 days after hatching, but probably in quantities too small to compensate for the vast numbers that die near the thousands of intakes of irrigation water. Only two Iranian units, together releasing $4-7 \times 10^6$ fingerlings a year, seem to work well. Yet poaching, which increased dramatically after the disintegration of the Soviet Union, is difficult to control, because the riverine states do not agree on the legal status (lake or sea) of the Caspian.

Caviar production has thus gone from bad to worse (13,300 tonnes in 1990 to 2,100 tonnes in 1994). A particularly alarming signal came last April, when the Iranian Fisheries Organisation at Nowshahr tried to demonstrate to us how it catches spawners for its hatcheries. To its embarrassment, it caught none. Other stations fared no better.

A decisive blow to the fisheries might come if *Mnemiopsis leidyi*, a voracious West Atlantic ctenophore, reaches the Caspian. The Lenin Canal between the Don and the Volga allows not only navigation between the Caspian and the Black Sea, but also faunal interchange that has already driven several Caspian endemic species to extinction and caused unwelcome hybridization between Black Sea and Caspian Sea sturgeon. *M. leidyi* reached the Black Sea in the 1980s,

destroying the local pelagic food chain, with the ensuing collapse of fisheries. It could be transported to the Caspian in ballast water: because traffic through the canal is heavy, with no regulation on ballast water, the introduction of *M. leidyi* seems only a matter of time.

What, if any, are the remedies? Like the Californian condor, the sturgeon's only chance of survival may be in captivity. Fortunately, rearing technology is available to keep this Caspian symbol alive in artificial lakes until its original home is ready to welcome it again. Meanwhile, stringent control of poaching and ballast water is important, as well as pollution abatement around the basin. But Turkmenistan is now extending the 1,200-km Karakumskiy Canal from the Amu-Darya river to the Atrek basin: drainage water contaminated with pesticides sprayed on cottonfields in north and east Turkmenistan may soon enter the Caspian via the Atrek as a new source of pollution. So, ironically, yet more water that would normally flow into the drying Aral will end up in an already overflowing Caspian.

What if the Caspian continues to rise? The phenomenon is not yet understood, although it has been known since antiquity that there is a cyclical pattern of rise and fall. There are at least a dozen possible explanations, in two broad categories: tectonic forcing and climate change, the main influence being that of the North Atlantic oscillation on precipitation in the

Volga basin². In the final analysis, Caspian people must learn to live with a fluctuating lake level.

Some relief has been provided by the reopening in 1992 of the giant evaporation pan of Kara Bogaz Gol on the east side of the Caspian. The pan and the sea were cut off from one another by a dam in 1980, built when the last recession was at its worst. The pan may accumulate up to 25 km³ of water, and in three years its refilling has prevented the Caspian from rising a further 35 cm. Thousands of smaller depressions scattered along the east coast could also be artificially flooded — although not before they are first tested for toxic deposits.

But the most perplexing 'solution', proposed by the Committee of Water Resources of Kazakhstan, is the digging of a canal between the Caspian and the Aral. The canal would be about 500 km long and involve raising the flow of a sizeable river by about 80 metres. The committee claims that only three pumping stations are needed, and the energy required would be supplied by the Gur'evskaya power plant. But it is unclear whether this 350-MW plant is adequate, and inevitably there will be ecological objections. Besides, the Caspian will probably start receding before the capital is raised and the canal dug.

Conservation will require an enormous investment. The littoral governments have been encouraged to work together and submit a funding proposal to the Global Environmental Facility, a US\$2.2 billion fund, administered jointly by the World Bank and the United Nations Development and Environment Programmes, and aimed at addressing environmental problems of planetary importance. If this is successful, several private donors may also contribute millions of dollars (although the Kazakhcaspihelf consortium, a group of mainly Western oil companies currently carrying out a seismic survey of the northern Caspian in search of more oil, has just withdrawn its support). Nor can the European Union remain indifferent. Indeed, the amount of money ultimately required to save the Caspian may be so large — several billion dollars for the clean-up of the Volga valley alone, for instance — that the World Bank, the European Bank for Reconstruction and Development and, of course, the littoral governments themselves may be the only credible investors available. □

Henri Dumont is at the Institute of Animal Ecology, State University of Gent, K. I. Ledeganckstraat 35, B-9000 Gent, Belgium.

Fall and rise of the world's largest lake

The Caspian Sea is the largest (384,400 km²) and most voluminous (about 78,700 km³) inland body of water on Earth¹. It lies in an elongated depression between the European and Asian continental plates, its surface well below sea level, with a maximum depth of 980 metres in the south and a shallow northern half averaging only 5.2 metres. The sea extends more than 1,000 km from north to south, and is bounded by deserts in the north and east, and by grasslands and forests in the west and south. It is an important source of oil and caviar.

The Caspian's low salinity is due to freshwater input. The Volga river contributes up to 82 per cent of the inflow, with the rest supplied by some 130 other rivers, principally the Ural, Kura and Atrek. At 3,530 km long, the Volga is Europe's largest river, housing half of Russia's population along its length and supplying a quarter of Russia's industrial and agricultural output. Between 1937 and 1981, it was transformed into a chain of reservoirs, designed mainly for the electrification of the Russian countryside and its

large-scale industrial and agriculture development.

The Caspian varies in level depending on evaporation and river input. It made the news in the 1960s when it started receding from its shores (from 126 metres in 1930 to 129 metres in 1977). The decrease was widely held to be due to the filling of the reservoirs along the Volga, but in fact began before these were built. In the 1960s, Soviet scientists advocated maintaining the lake at a minimum level by diversion of water from northern rivers, thus starting the Soviet Union's first major ecological debate.

The controversy ended abruptly in 1978 when, unexpectedly but rapidly, the Caspian started to rise. Today its level is 126.5 metres and rising. Typically lacking in ecological memory, people moved into the newly emerged lands, developing agriculture (as in the Kura and Volga plains), exploiting oil reserves (as in the Apsheron and Cheleken peninsulas) and establishing townships (as along the Iranian coast, where population density is greatest). All these regions are now flooded.

H. D.

1. Kosarev, A. N. & Yablonskaya, E. A. *The Caspian Sea* (SPB Academic, The Netherlands, 1994).

2. Rodionov, S. N. *Global and Regional Climate Interaction: The Caspian Sea Experience* (Kluwer, The Netherlands, 1994).

What remains, however improbable ...

The interpretation of the fossil remains of long-extinct creatures with no close modern relatives poses problems familiar to fog-bound Victorians — and modern palaeontologists.

WHEN you have eliminated the impossible (said Sherlock Holmes), what remains, however improbable, must be the truth. Holmes was as logical as he was lonely — sometimes, it is as if Conan Doyle's fictional Victorian detective represents the only rational beacon (apart from the faithful Dr Watson, of course) in a world in which fog-bound phantoms ever haunt the darker ends of the gas-lit streets, a district of opium dens, dogs that do not bark, the likes of Stevenson's Mr Edward Hyde, and virtually anybody from a Poe story. Palaeontologists know this neighbourhood well. They are quite used to working with fossils that defy interpretation, the remains of animals that somehow, even after everything else has been eliminated, seem to bend the pillars of the probable.

Such are the facts in the strange case of three tiny teeth from a well-known mammal fauna from the Early Cretaceous of Morocco, interpreted by their discoverer, Denise Sigogneau-Russell of the Institut de Paléontologie in Paris, France, as belonging to triconodonts, members of a long-vanished order of mammals (*Acta Palaeontologica Polonica* 40 (2), 149–162; 1995). One of the teeth, although with unusual, wickedly pointed cusps, resembles known triconodonts closely enough for it to be placed in that order with fair confidence, although it may represent a hitherto unknown family. Sigogneau-Russell names it *Dyskritodon amazighi*: the generic name — 'a tooth of uncertain placement' — reflects her understandable misgivings about her creation.

The other two teeth are far, far stranger. For, despite their each having paired roots (usually a good sign of a mammalian molar), they are thin and blade-like, rather than squarish and blocky. The cusps — three large ones with a smaller at the trailing edge — are deep, sharp and pointed. The whole affair looks less like a tooth than a miniature stone-age bone harpoon. They are clearly different from the sole tooth of *Dyskritodon*, but similar to each other, and Sigogneau-Russell corrals them both into the same species, *Ichthyoconodon jaworowskorum*. The generic name reflects a double meaning — these teeth look as if they come from a fish-eater, but they could also have come from a fish. Their phylogenetic placement, somewhere among the triconodonts, is even less certain than that of *Dyskritodon*.

Sigogneau-Russell spent much time

eliminating the impossible before settling on the Triconodonta as the home for these improbable teeth. Even scolecodonts — the microfossil jaws of worms — did not evade her scrutiny. Among vertebrates, her first call was among the fish, sharks in particular. Sharks (especially fossil ones) are unmatched in the novelty dentition department, but the presence of paired roots seemed to rule out *Ichthyoconodon* from their catholic membership. Again, the paired roots bar these barbs from the squamates (lizards, snakes — and other things). Dinosaurs (and they would have to be very petite dinosaurs) are unknown to have such teeth. Even pterosaurs, notorious for their *outré* orthodontics, cannot admit *Ichthyoconodon* to their be-fanged ranks. Among the mammals, Sigogneau-Russell examined the multituberculates. These Mesozoic mammals (extinct by the Eocene) sometimes have very specialized, multi-cusped teeth, but (among other things) the roots are not as well-separated as in *Ichthyoconodon*. So, triconodonts they are, at least for the time being.

The bulging casebook of masticatory mysteries shows that few things have such a memorable ability to sow confusion and dissent as teeth. Two years ago, in this column (*Nature* 360, 529; 1992), I discussed the debate about three teeth in a fragmentary jawbone from the Palaeocene of Canada, claimed by R. C. Fox *et al.* (*Nature* 358, 233–235; 1992) to have belonged to a member of a group of mammal-like reptiles otherwise extinct since the Jurassic. That claim has since been disputed, but until more of the skeleton of *Chronoperates paradoxus* is described, the mystery must remain unsolved.

The problem, in part, is that the design of teeth — at the sharp end, as it were, of the sustenance of the entire animal — may say more about present function than past history. The naming of *Ichthyoconodon* as a fish-eater is wise, as fish-eaters of all kinds tend to have simple, pointed teeth that reveal little about ancestry. The teeth of toothed whales, for example, are far from the multi-cusped molars of their terrestrial progenitors. To choose another example, the rodents arguably owe their success as the most speciose mammalian order to their specialized dentition — the result is that rodent teeth all look much of a muchness to the uninitiated, making rodent systematics a formidable task. This problem makes the assignment of *Ichthyoconodon* to the triconodonts (or to

anything else) a brave one.

A wider problem concerns the way in which palaeontologists, as comparative zoologists, interpret their fossils. One can make sense of a new specimen only by comparing it with material whose interpretation is already secure. Pleistocene palaeontologists have it easy, as most Pleistocene fossil forms have relatives alive today. Matters become harder as one goes back in time, to encounter forms that bear no close resemblance to anything now living. Such creatures hover at the dark end of the street. One thinks of the peculiar faunas of the Cambrian 'explosion' (see the report by Chen *et al.* on page 720 of this issue, with the accompanying News and Views by Stephen Jay Gould on page 681).

The twin problems of function and fossil interpretation collide with the conodonts. These phosphatic, tooth-like microfossils are sufficiently abundant in many Palaeozoic strata to be biostratigraphically informative, yet, until recently, the nature of the conodont-bearing animal was something to be guessed at. (Imagine the task of an alien palaeontologist of the future, trying to deduce the human frame from several million sets of dentures.) The discovery of the first body-fossils of conodont animals (D. E. G. Briggs *et al.*, *Lethaia* 16, 1–14; 1983) did not immediately solve the problem, the animals being interpreted as chordates, chaetognaths, or molluscs with highly specialized shells. It now seems clear that conodonts were feeding structures of elongate, fish-like creatures closely related to vertebrates (P. Janvier, *Nature* 374, 761–762; 1995), although this interpretation has yet to receive universal approbation.

Swaddled in disorienting mist, palaeontologists live in a world beyond holmesian logic, perhaps closer to that imagined by another eminent Victorian: a landscape of dark frontiers peopled by borogoves, momewraths, mock turtles, and royalty quite happy to believe as many as six impossible things before breakfast. **Henry Gee**

Correction

John Maddox, in "Big Bang not yet dead but in decline" (*Nature* 377, 99; 1995), referred to results obtained by M. J. Pierce *et al.* using the Keck telescope. In fact their results were obtained at the Canada-France-Hawaii telescope, which is also in Hawaii. □

Tuning retinal circuits

Peter Sterling

WHILE studying the retina more than 100 years ago, Santiago Ramon y Cajal noted that deposits of silver dichromate completely filled the arborization of a single neuron but stopped at the cell boundary. He concluded that contiguous neurons are discrete and that signals between them must cross an extracellular gap, now known as a chemical synapse. He vigorously defended this idea against 'reticularism', the view that neurons form continuous networks, with his penetrating observations and towering polemics, and by 1935 reticularism was apparently crushed¹.

But we now know that the reticularists were also right: retinal neurons can couple with one another by means of clusters of fine intercellular channels called gap junctions² and it seems that this coupling is crucial for retinal function. New results described on page 734 of this issue by Mills and Massey³ offer a deeper insight into reticularism with the demonstration that these gap junctions differ in pore size at different points in a circuit and that they are regulated by different second messengers.

The authors injected tracers of different molecular weights into an AII amacrine cell of the rabbit retina (maintained *in vitro*) and followed their spread into adjacent neurons to which the AII is known to be coupled. As expected⁴, the smaller tracer spread into the neighbouring AII cells and cone bipolar cells. A larger tracer with the same charge also spread readily into AII cells, but penetrated poorly into cone

bipolars (see Figs 2 and 3 of Mills and Massey's paper³). High concentrations of cyclic AMP were already known to curtail tracer spread to AII cells⁴, and the authors found that cyclic GMP had a similar effect on tracer spread to cone bipolar cells (see their Fig. 4).

This discovery of a difference in pore size and in second messengers raises a host of questions that may eventually link the architecture of microcircuits to that of their computationally important molecules. To grasp the key issues requires some knowledge of the overall circuitry in which the AII cell is a critical link (see figure).

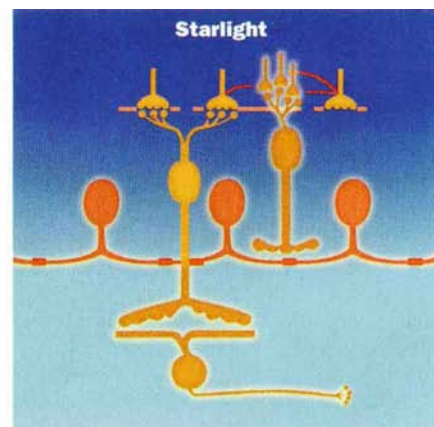
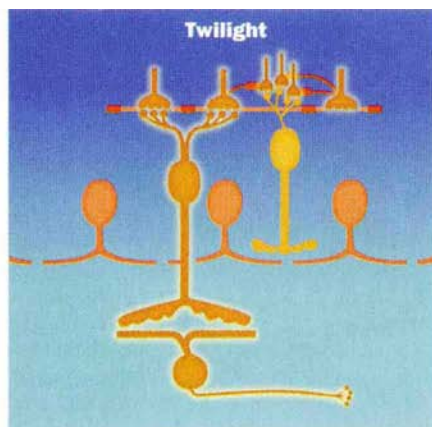
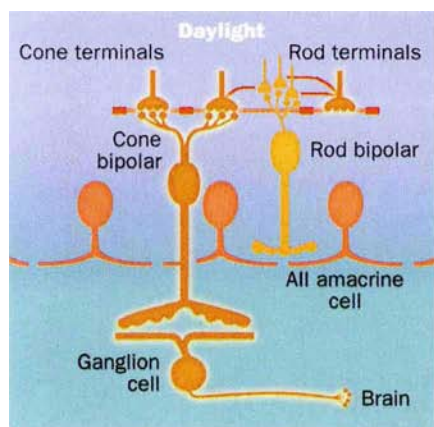
Neural signals for daylight and starlight are transduced by different photoreceptors (cones and rods respectively) and traverse the retina over different circuits — until the last stage, which is transmission to the ganglion cell⁵. This stage requires 10^2 – 10^3 chemical synapses (because each synapse operates stochastically), and these occupy most of the ganglion cell's dendritic surface⁶. So, to provide separate sets of synapses would require either fewer synapses per circuit (compromising signal quality) or more space.

An alternative design has been adopted universally by mammals: cone bipolar cells contribute one full set of excitatory chemical synapses to the ganglion cell, and rod bipolar cells 'borrow' them at night by chemically exciting the AII cell, which through extensive gap junctions⁵ passes depolarizing (excitatory) ionic currents into the cone bipolar synapses.

Direct coupling avoids both delay and additional stochastic fluctuation and functions as a simple manoeuvre to transfer the starlight signal to the ganglion cell.

But there is a catch. In daylight, cone-evoked depolarizations should spread backwards from the cone bipolar synapse into the AII cell and then laterally, as a result of coupling between neighbouring AII cells. This would expand the summation area of the individual ganglion cell well beyond its dendritic field and thus degrade the spatial detail carried by the ganglion cell array. So, once the neural connections were recognized as a 'circuit', it was predicted that the AII-cone bipolar junction should uncouple in daylight⁷. The demonstration by Mills and Massey that coupling at this junction is indeed reversible is gratifying and strengthens our interpretation of retinal circuitry.

The coupling between AII cells addresses an entirely different problem. In one second of starlight, fewer than one rod in a thousand transduces a photon. Therefore the several thousand rods wired to a ganglion cell carry mostly noise, which potentially obscures the rare light events⁵. But photons reflected from an object in the visual scene, although thinly spread, are correlated. Coupling spreads the correlated signal to many AII cells, reducing signal amplitude but reducing the uncorrelated noise even more. This manoeuvre improves the signal-to-noise ratio in the AII cell, which can then remove residual noise by using voltage-sensitive sodium channels to set a threshold^{8–10}. Coupling in the AII network may increase gradually with nightfall in order to optimize spatial averaging to the



To convey the full range of environmental intensities requires three different but partially overlapping neural circuits. The mammalian retina accomplishes this at no extra cost in space or noise by using four sets of gap junctions (red), at least three of which are regulated. In daylight, cone terminals excite (depolarize) cone bipolar cells that chemically excite the ganglion cell. This last excitatory stage occupies most of the ganglion cell's dendritic surface (some space is reserved for inhibitory synapses). Gap junctions couple the cone terminals strongly enough to improve signal-to-noise ratio, but not so much as to degrade acuity. In twilight, cones are less active, but the 50 rods surrounding each cone couple to it via their terminals. In effect, rods 'parasitize'

the cone terminals and thus the rest of the cone bipolar circuit. In starlight, only one rod per thousand transduces a photon over one second, so the rod-cone junction conveys mostly noise. Consequently, this junction uncouples, and rod terminals excite rod bipolar cells to excite AII amacrine cells chemically. The AII cell couples to the cone bipolar axon, in effect parasitizing the cone bipolar-to-ganglion cell synapses. Gap junctions couple AII cells strongly enough to spread current widely and thus enlarge the ganglion cell's summation area well beyond its dendritic tree. This would improve signal-to-noise ratio in the dimmest light, but degrade acuity in brighter light, so this junction is also regulated.

declining quality of the retinal image.

Why are different messengers used to regulate the forward signal spread (AII to cone bipolar) and the lateral signal spread (between AII cells)? Probably because the forward junction needs to turn on fully at the transition from twilight to starlight (see figure), whereas the lateral junctions turn on gradually over a thousand-fold intensity range. It may be no accident that the forward junction is uncoupled by cGMP because light raises cGMP in the cone bipolar cell. A key next step will be to determine whether the cone terminal's synaptic modulation of cGMP does the trick or whether, as suggested by Mills and Massey³, nitric oxide might be required. The rod-to-cone junction, which operates in twilight, uncouples¹¹ (see figure) just as the AII-to-cone bipolar junction couples, and could well require yet another second messenger.

Most puzzling, perhaps, is why different junctions use different pore sizes. If current spread were the only issue, to vary the number of channels per junction might be simpler. Thus, pore size might reflect the need to exchange certain molecules between AII cells but not between AII and cone bipolar cells. If so, there is no hint as to what they might be, and we should bear in mind that different pore size might merely accompany, as an epiphenomenon, the requirement for differential regulation. The single channels form as hexamers of the protein connexin, of which there are many isoforms¹². The next step would be to identify specifically which connexins are expressed on each side of these junctions. This will take reticularism down to where neuro-nism has already arrived, the molecular level. □

Peter Sterling is in the Department of Neuroscience, University of Pennsylvania Medical Center, 123 Anatomy-Chemistry Building, 36th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6058, USA.

1. Ramon y Cajal, S. *Neuron Theory or Reticular Theory* (1935); translation by Purkiss, M. U. & Fox, C. A. (Consejo Superior de Investigaciones Científicas, Instituto Ramon y Cajal Madrid, 1954).
2. Vaney, D. I. *Progr. Ret. Res.* **13**, 301–355 (1994).
3. Mills, S. L. & Massey, S. C. *Nature* **377**, 734–737 (1995).
4. Hampson, E. C. G. M., Vaney, D. I. & Weiler, R. J. *Neurosci.* **12**, 4911–4922 (1992).
5. Sterling, P., Freed, M. A. & Smith, R. G. *J. Neurosci.* **8**, 623–642 (1988).
6. Kier, C., Buchsbaum, G. & Sterling, P. *J. Neurosci.* (in the press).
7. Sterling, P., Cohen, E., Freed, M. A. & Smith, R. G. *Neurosci. Res. (suppl.)* **6**, S269–S285 (1987).
8. Nelson, R. J. *Neurophysiol.* **47**, 928–947 (1982).
9. Boos, R., Schneider, H. & Wässle, H. *J. Neurosci.* **13**, 2874–2888 (1993).
10. Smith, R. G. & Vardi, N. *Vis. Neurosci.* **12**, 851–860 (1995).
11. Schneeweis, D. M., & Schnapf, J. L. *Science* **268**, 1053–1056 (1995).
12. Bennett, M. V. L., Bargiello, T. A., Spray, D. C., Hertzberg, E. & Saez, J. C. *Neuron* **6**, 305–320 (1991).

Magnetic grains in GaAs

Tom M. Giebultowicz and Valerie Nunez

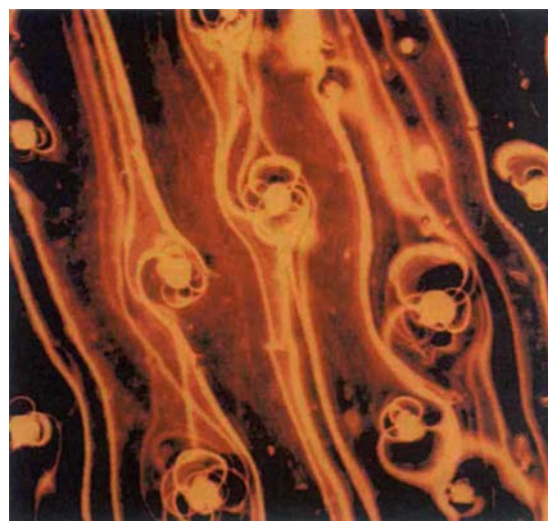
As the fabrication of structures such as quantum wells in semiconductors has become commonplace, a whole world of new ultra-compact electronic devices has been opened up. If discrete magnetic components could be incorporated into such structures, their technological potential would be increased still further; but, so far, combining magnetic compounds with semiconducting matrices to produce sizeable magnetic effects has proved difficult. Now, as they describe on page 707 of this issue, Jing Shi and colleagues¹ have developed a new method of incorporating minute ferromagnetic grains into a semiconducting matrix.

The host material, a thin film of gallium arsenide, is first implanted with high-energy Mn ions. Subsequent thermal treatment activates a crystallization process, resulting in *in situ* formation of submicrometre GaMn grains. From sophisticated techniques that allow one to examine individual grains, Shi *et al.* found that the crystallographic structure of the insets was different from all other known bulk GaMn modifications. In particular, the data revealed five-fold symmetry, the signature of quasicrystallinity. This is intriguing in itself, but the more important point is that the grains were ferromagnetic — good news indeed for semiconductor physicists.

Artificially structured systems prepared by 'atomic engineering' methods are currently much in vogue with condensed matter physicists and material scientists. These systems may be granular solids of various forms (grains embedded in a solid matrix or contained in 'nutshells' made of another material) or multilayers prepared on flat surfaces by depositing layers of two or more different materials in combination. Such heterostructures show physical effects not seen in bulk solids, many of which may prove useful for building new generations of miniature electronic or optoelectronic devices. Perhaps the best-known example is that multilayers composed of different semiconducting materials are practical realizations of quantum-mechanical potential wells. By changing layer thicknesses and compositions, one can control well characteristics such as depth and shape, and thus manipulate the quantum states of entrapped electrons to obtain desired patterns of energy levels. The principle has been used for the design of a blue-green solid state

laser² which could quadruple the capacity of compact discs.

Systems fabricated from combinations of magnetic and non-magnetic materials are an important class of artificially structured solids. The mobile and valence electrons that determine the electronic properties of such systems interact with the magnetic atoms, further modifying their quantum states. This gives rise to a variety of new effects, many of which seem to be strongly sensitive to external magnetic fields, and might therefore be exploited for constructing new field-



Microscopic crystals of ferromagnetic GaMn (shown in this transmission electron micrograph as bright grains typically 200 nm in diameter) in a semiconductor matrix¹.

Jing Shi *et al.*

tunable electronic sensors and devices. One effect that is currently attracting much attention is giant magnetoresistance, an unusually strong dependence of the electric conductivity on external magnetic field, which is seen in certain metallic granular solids as well as in multi-layered structures.

Strong interest in semiconducting systems containing magnetic implants was first aroused in the mid-1970s following the discovery that substituting magnetic ions for a small fraction (say, 5 per cent) of the metallic cations in a semiconducting material leads to profound changes in the material's optical, magneto-optical and conducting properties. For example, such alloys (usually referred to as 'semi-magnetic semiconductors') show an unusually strong Faraday effect — in a magnetic field, they rotate the polarization vector of transmitted light by an amount proportional to the applied field intensity. This effect can be used for building light modulators as well as 'light valves' for fibre-optic technology.

Currently, research on epitaxial multi-

layered structures based on these semiconductor materials is focused primarily on two families of systems. The first is multilayers and heterostructures derived from the II–VI semiconducting materials³ (such as CdTe and ZnSe), in which the magnetic constituents are isostructural Mn compounds (MnTe, MnSe). The other is based on IV–VI semiconductors such as PbTe and rare-earth magnetic chalcogenides⁴ such as EuTe. Some of the unusual magnetic properties of these systems are of great fundamental interest, for instance the formation of ‘spin superlattices’ in which the electronic spin-up and spin-down states become spatially separated^{5,6}. Also, results⁷ from digital heterostructures (quantum wells in which the electronic states are tailored through controlled distribution of two-dimensional magnetic layers) confirm that multilayered magnetic semiconductor structures offer considerable technological opportunities: the Faraday effects are two orders of magnitude stronger than in bulk semimagnetic semiconductors, and the dynamics of the spin-dependent electronic processes are extraordinarily fast, and should thus be capable of handling much faster signal transmissions than in current state-of-the-art communications.

As yet, though, technologists develop-

ing new magnetic semiconductor quantum-well structures have been struggling to overcome two challenges. The first is fabrication of magnetic semiconductor structures with ferromagnetic components. By a caprice of nature, magnetic compounds that can be combined with the most widely used semiconducting materials are almost exclusively antiferromagnets, with no net magnetic moment. Systems with ferromagnetic components would be far more sensitive to external magnetic fields, further enhancing their technological value. The second challenge is to grow magnetic semiconducting structures based on the III–V compounds (such as GaAs, InSb) which are particularly advantageous in optoelectronic applications.

Set against this background, we can see that the work of Shi *et al.* represents a great step towards overcoming both these obstacles. The host material used in their work, GaAs, is representative of the III–V family; and the micrograins implanted are ferromagnetic. Furthermore, the Curie temperature for this GaMn modification is well above room temperature.

Of course, problems remain. At present, the new technique produces a random distribution of GaMn grains. It seems unlikely that it can be used for preparing continuous layers, as in systems

fabricated by molecular beam epitaxy. But the recent successes of focused ion beam techniques⁸ suggest that the preparation of patterned ferromagnetic arrays is possible. It should be noted that the newest trend in research on magnetic semiconductor multilayers is to study systems with modulated structures within the layers — for example, regular patterns of magnetic dots or ‘wires’ buried in the material. Once more control is gained over the process of GaMn precipitation, the method described by Shi *et al.* may become an ideal tool for fabricating such specimens. □

Tom M. Giebultowicz is in the Department of Physics, Oregon State University, Corvallis, Oregon 97331, USA. Valerie Nunez is in the Reactor Radiation Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA.

1. Shi, J. *et al.* *Nature* **377**, 707–710 (1995).
2. Luo, H. & Furdyna, J. K. *Semiconductor Science and Technology* **10**, 1041 (1995).
3. *Diluted Magnetic Semiconductors* (Semiconductors and Semimetals, vol. 25) (eds Furdyna, J. K. & Kossut, J.) (Academic, Boston, 1988).
4. Yuan, S., Springholz, G., Bauer, G. & Kriechbaum, M. *Phys. Rev. B* **49**, 5476 (1994).
5. Chou, W. C., Petrou, A., Warnock, J. & Jonker, B. T. *Phys. Rev. Lett.* **67**, 3820 (1991).
6. Dai, N. *et al.* *Phys. Rev. Lett.* **67**, 3824 (1991).
7. Crooker, S. A. *et al.* *Phys. Rev. Lett.* **75**, 505 (1995).
8. Kikkawa, J. M. *et al.* *Phys. Rev. B* **50**, 2003–2006 (1994).

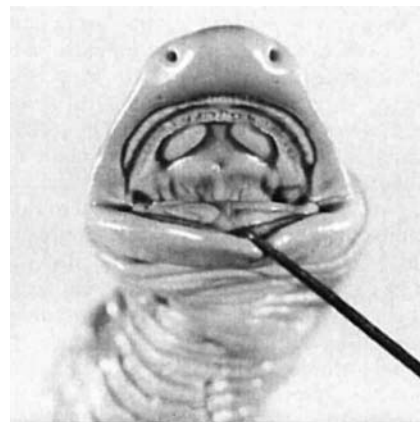
Don't hold your breath

EXTINCTION may already have claimed an otherwise revolutionary new form of vertebrate life: an erstwhile land animal with relatively huge, gaping jaws (pictured here), that lacks limbs, lungs, pulmonary circulation and even internal nostrils. The species concerned is a caecilian, *Typhlonectes eiselti*, and according to Ronald A. Nussbaum and Mark Wilkinson (*Proc. R. Soc. Lond. B* **261**, 331–335; 1995), its unusual features could mark the start of an adaptive radiation of secondarily aquatic vertebrates.

Caecilians constitute the least known of the three living orders of amphibian. All extant species are limbless, although this condition is secondary (see *Nature* **365**, 246–250; 1993). Their tropical distribution and mainly burrowing or aquatic lifestyle has ensured their obscurity: many of the 160 or so species recognized are known from just one or two specimens. *Typhlonectes eiselti* is known only from a single preserved specimen in the Vienna Museum of Natural History, collected somewhere in South America, no later than the 1920s, and possibly much earlier.

Amphibians ventilate by pumping air into their lungs by way of the mouth cavity and internal nostrils, or choanae.

The fact that the choanae in *T. eiselti* are sealed, a condition unique in tetrapods, caught Nussbaum and Wilkinson's attention, and dissection revealed a complete lack of lungs and supporting pulmonary circulation. Freedom from the necessity of buccal venti-



lation has allowed an extraordinary redesign of the skull. It is larger and looser than most caecilian skulls, with distinctly aberrant musculature, and a jaw articulation placed well back that allows for a very wide gape. This complex of features — the obligate aquatic lifestyle and redesigned skull — has the

potential to free *T. eiselti* from the usual caecilian burrowing habit, and permit the foundation of an entirely new type of aquatic vertebrate (one thinks of *Megachasma*, the recently described genus of deep-sea shark).

What of lunglessness? Many amphibians use the skin as a gas-exchange surface, and some aquatic salamanders either retain lungs as buoyancy aids, or dispense with them altogether. Lunglessness tends to be associated with small size, low temperatures (and thus low metabolic rate) and high ambient oxygen concentration, such as one might find in fast-flowing mountain streams. *Typhlonectes eiselti* is 725 mm long, and is by far the largest lungless vertebrate known, something that would limit its environmental choices. Although the details of where it was collected are unknown, Nussbaum and Wilkinson suggest that it might inhabit cool, fast-moving streams in montane rainforests.

The fact that such habitats are currently threatened by human activity — and that no other specimens of this species have been collected for at least 70 years — suggests that this bold experiment in vertebrate design might have met an untimely end. H. G.

The making of metal deposits

Randolph A. Koski

ON page 713 of this issue¹, an eagerly awaited paper by Humphris *et al.* describes the initial results of drilling into an actively forming massive sulphide deposit on the Mid-Atlantic Ridge, and through into the underlying bedrock. This is something of a technical achievement for the Ocean Drilling Program (ODP) and it provides, for the first time, a glimpse of the processes operating within such a deposit. The resulting information is of added significance because many sources of metals on land originally formed in this way.

High-temperature hydrothermal vents ('black smokers') at oceanic ridges are important agents of mass and fluid transfer between the oceanic crust and the ocean². Although much of the iron, copper, zinc and other metals mobilized from the underlying oceanic crust during fluid-rock interaction is discharged into the ocean as black smoke, some metals are precipitated as sulphide deposits at the vent site and within underlying volcanic rocks. These sea-floor mineral deposits are the modern-day equivalents of the economically valuable massive sulphide deposits preserved as parts of ophiolites on land, in Cyprus and Oman for example.

The discovery of sea-floor hydrothermal systems has been a boon to ore-deposit research. Especially relevant are sea-floor deposits that have a size (millions of tons) commensurate with their ancient counterparts, perhaps the best-known being the Transatlantic Geotraverse (TAG) sulphide mound on the Mid-Atlantic Ridge. This is the mound that was drilled by the ODP, and the subject of Humphris and colleagues' paper.

It was the discovery, in the late 1970s, of sulphide-forming hydrothermal vents on the East Pacific Rise that confirmed the submarine origin of volcanic-associated massive sulphide (VMS) deposits preserved in ophiolite and island-arc terranes of many orogenic belts³. The details vary, but a general picture of VMS deposits is that they consist of a concordant lens-shaped body of massive sulphide (60 per cent or more of sulphide minerals) overlying a discordant pipe- or funnel-shaped zone of altered volcanic substrate known as the stockwork^{3,4}. Models for the formation of VMS deposits have relied heavily on systematic geochemical and mineralogical investigations of the ores and their wallrock. For example, the range of sulphur isotope ratios in samples of massive sulphide ore could best be explained if the dissolved sulphide in hydrothermal fluids came in part from the underlying volcanic rocks and in part from the inorganic reduction of seawater sulphate³. Analyses

of mineral textures led to the realization that metal zonation in VMS deposits was related to temperature-dependent replacement reactions within maturing sulphide mounds⁵; and secondary mineral compositions and fluid inclusion data from stockworks below massive sulphides have been used to set constraints for the composition and temperature of ore fluids and to estimate fluid/rock ratios⁶.

With few exceptions, however, descriptive studies of VMS deposits yield an incomplete record of their genesis. During their formation on the ocean floor, recrystallization and replacement processes will have continually modified the active

hydrothermal vents (some 6–200 per cent of the seawater value) that can only be attributed to fluid separation at depth into vapour and brine phases^{7,8}.

Among the wide variety of sea-floor hydrothermal deposits now known, attention has centred on the most pristine and accessible constructions — the black smoker chimneys growing on the tops of mounds. Chimneys are essentially mini-deposits, and are convenient places to learn about controls exerted on precipitation of sulphides and sulphates, and on ore-deposit formation, by the steep temperature and chemical gradients set up within the zone where hydrothermal fluids and seawater mix. But the problem of scale arises. Only relatively small chimneys (a few centimetres to a few metres in height) have been available for study, generally in fragments retrieved by dredging

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Sea life around the orange iron oxide and sulphide surface of the TAG mound.

deposits, and it is unlikely that any of the early mineralization survives. Once formed, the sulphide deposits are degraded by mass wasting and oxidation. Through time, deposits are further modified by burial, metamorphism, deformation and, in some cases, subaerial weathering.

Given these considerations, the discovery of hot springs and massive sulphides along the crests of the East Pacific Rise and Mid-Atlantic Ridge has paid big dividends in the effort to decipher VMS deposits. Perhaps the most valuable data have been gained from hundreds of samples of vent fluids (modern ore fluids) painstakingly collected from discharge sites on chimneys and mounds. On bare-rock ridges such as TAG, endmember hydrothermal fluids with high temperatures and salinities, and low pH, carry much greater loads of dissolved metals, including copper and zinc, than does seawater⁷. These are potent ore solutions. Of additional significance to the transport of metals is the variation in salinities among

or submersible, and the applicability of such data to the formation of massive sulphide deposits is limited. With an estimated mass of about 3 million tons of massive sulphide and an elliptical form^{1,9}, the TAG mound has a bulk and shape comparable to many VMS deposits on land.

So how did the TAG deposit grow to this size? Longevity is certainly one factor, TAG being over 20,000 years old¹⁰. But drilling deep within the mound structure, as described by Humphris *et al.*, has revealed new clues regarding the actual accretionary process. It turns out that the mound contains extensive breccia deposits, indicating that both high- and low-temperature chimneys atop the mound have repeatedly grown and collapsed to produce sulphide fragments that, along with fragments of the underlying volcanics, are gradually cemented and replaced during subsequent reactions with hydrothermal fluids percolating through the porous infrastructure. During replacement, more soluble metals such as zinc

are remobilized from chimney fragments to the outer reaches of the mound contributing to large-scale metal zonation within the deposit¹¹. This cycle of reworking and refining must repeat itself many times as the mound increases in size.

Drilling of TAG also revealed large veins of anhydrite, a mineral that is relatively scarce in ancient deposits. Although the involvement of anhydrite in chimney building is well established¹², it must also play a larger role in the mound building (and perhaps the mound wasting) process than has been realized. Its precipitation requires entrainment of sea water, a process that must also cool the mound interior. The reduction in porosity related to deposition of anhydrite within the mound will tend to localize and focus the upward flow of high-temperature fluids. This mechanism may be critical for mound development as it controls the location of chimney complexes. Anhydrite is uncommon in ancient deposits because of its retrograde solubility with respect to temperature. So, as mound temperatures wane, anhydrite quickly redissolves into sea water, causing increased porosity and instability within the structure.

The ODP investigations of TAG and hydrothermal sites at sediment-covered spreading axes such as Middle Valley on northern Juan de Fuca Ridge¹³ have greatly improved our knowledge of how volcanogenic ore deposits are formed. More is to come. Deeper penetration at TAG and other hydrothermal systems will add information about the three-dimensional architecture of sub-seafloor circulation and mineralization; and the eventual drilling of vent areas in other tectonic settings, such as volcanic arcs and backarcs, will tell us more about VMS deposits formed in those palaeoenvironments. Such drilling programmes are highly demanding technically. But as these latest results demonstrate, the scientific dividends can be great. □

Randolph A. Koski is in the US Geological Survey, Branch of Pacific Marine Geology, 345 Middlefield Road, MS 999, Menlo Park, California 94025, USA.

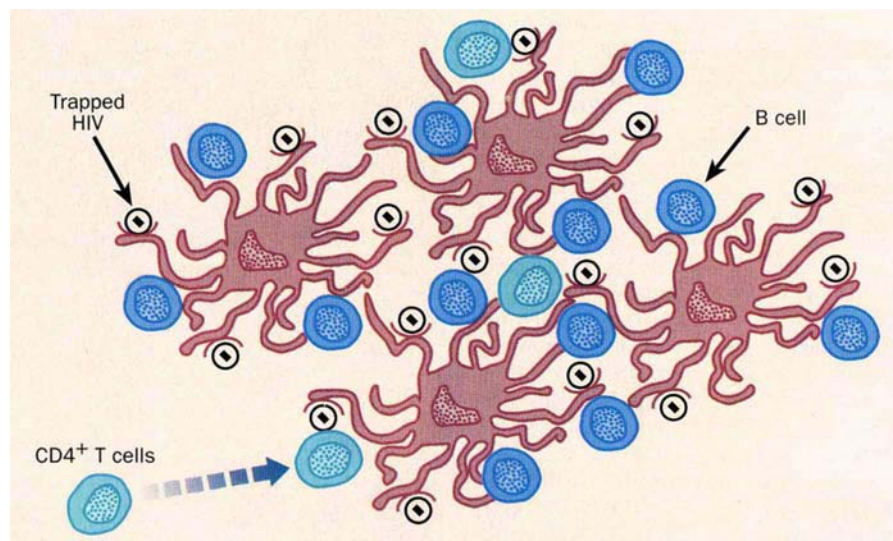
Trapped but still dangerous

Lewis K. Schrager and Anthony S. Fauci

A SINISTER survival tactic of HIV is revealed by Heath *et al.* on page 740 of this issue¹. It appears that when the virus is bound onto follicular dendritic cells (FDCs) in the germinal centres of lymphoid tissue, it remains highly infectious for CD4⁺ T cells even in the presence of antibodies that would otherwise neutralize the infectivity of the virus.

A variety of viral and host factors are implicated in the development of disease caused by HIV infection². Foremost among the host factors is lymphoid tissue, which is involved in the initiation and propagation of HIV's pathogenic effects^{3,4}. It has been known for some time

organs provide the critical microenvironment for driving HIV disease. In this model, HIV is widely disseminated to lymphoid tissue during the burst of viraemia that occurs in the acute phase of primary HIV infection. Immune cells are activated and germinal centres proliferate, their interwoven web of FDC processes forming a snare for antigens, in this case viral particles. A brisk immune response sharply curtails but virtually never completely suppresses virus replication. HIV is unique among viruses in that it continues to replicate for years, and hence viruses go on being trapped until the microenvironment of the lymphoid tissue



The web-like arrangement of follicular dendritic cells (FDCs) within the germinal centre of lymphoid tissue. Antibody- and complement-coated HIV virions are attached to the processes of the FDCs. CD4⁺ T cells that are travelling through the germinal centres to provide immunological help for the B cells responding to the virus are themselves susceptible to being infected by the infectious virions attached to the FDCs. (Modified from ref. 3 with permission.)

that the virus settles in lymphoid tissue during infection, where it takes up residence in the form of large numbers of latently infected cells⁵, and that it replicates throughout the course of HIV disease⁶, making lymphoid tissue the principal reservoir and site of replication of HIV. Clusters of extracellular virions become trapped on the surfaces of the web-like processes of FDCs, which permeate the germinal centres that form the focus of immune reactivity in lymphoid tissue (reviewed in ref. 3).

These findings, superimposed on the perpetual presence of the circulating virus⁷ and its extraordinarily high rate of turnover^{8,9}, provide a picture of dynamic viral disease even in the face of clinical latency. The potential importance of the extracellular trapping of viral particles on FDC processes was emphasized by Pantaleo *et al.*^{3,4}, who proposed that lymphoid

is ultimately destroyed⁶. Chronic and persistent HIV infection is thus established.

The primary purpose of antigen trapping on FDCs is to provide a sustained stimulus to the immune system and to maintain immunological memory, particularly B-cell memory. Although these are important benefits for the host, trapping gives HIV an advantage too³. On the one hand, trapping helps the induction and maintenance of a competent HIV-specific immune response, but on the other it could be harmful as a constant source of infection for CD4⁺ T cells, the chief target for HIV. This concern is justified in light of the fact that large numbers of CD4⁺ T cells reside in the paracortical region of lymphoid tissue and migrate into the B-cell-rich germinal centres to provide help for B cells that are mounting an immune response against the virus (see figure). The danger to these cells would be predi-

1. Humphris, S. E. *et al.* *Nature* **377**, 713–716 (1995).
2. Converse, D. R., Holland, H. D. & Edmond, J. M. *Earth planet. Sci. Lett.* **69**, 159–175 (1984).
3. Franklin, J. M., Lydon, J. W. & Sangster, D. F. *Econ. Geol.* **75**, 485–627 (1981).
4. Lydon, J. W. *Geosci. Can.* **11**, 195–202 (1984).
5. Eldridge, C. S., Barton, P. B. Jr & Ohmoto, H. *Econ. Geol. Mon.* **5**, 241–281 (1983).
6. Valsami, E., Cann, J. R. & Rassios, A. *Chem. Geol.* **114**, 235–266 (1994).
7. Von Damm, K. L. *Rev. Earth planet. Sci.* **18**, 173–204 (1990).
8. Von Damm, K. L. *et al.* *Nature* **375**, 47–50 (1995).
9. Thompson, G. *et al.* *Can. Mineralogist* **26**, 697–711 (1988).
10. Lalou, C. *et al.* *Earth planet. Sci. Lett.* **97**, 113–128 (1990).
11. Tivey, M. K. *et al.* *J. geophys. Res.* **100**, 12527–12555 (1995).
12. Haymon, R. M. *Nature* **301**, 695–698 (1983).
13. Mottl, M. J. *et al.* *Proc. ODP, Leg 139 Scientific Results* (1994).

cated on the assumption that the trapped virions are infectious. Certainly they should not be, for they are coated with antibody and complement, which should be more than enough to prevent them from infecting vulnerable CD4⁺ T cells.

What Heath *et al.* show is that, despite being coated with neutralizing antibody, virions that are trapped on the processes of FDCs are in fact still infectious. These ensnared virions can ambush any CD4⁺ T cells that migrate through the germinal centres. Interestingly enough, the concept of a virus in tissue being infectious even when it is covered with neutralizing antibody is not new: Albert Sabin described this phenomenon in 1935¹⁰.

There are other implications of these findings that relate to potential therapeutic strategies. It is conceivable that the processes of FDCs carrying clusters of viruses — even if they are surrounded by antibodies — may act as a protected sanctuary for infective virions, including resistant mutants whose presence will ultimately contribute to the failure of anti-retroviral therapy. It might be possible either to prevent the virus from becoming trapped in the first place or to remove the entangled virus from the FDC in an attempt to interrupt this vicious cycle. Selective agents have been shown to reduce the amount of antigen bound to FDCs in a mouse model¹¹.

If such an approach were successful in

purging trapped virus from lymphoid tissue, a source of *de novo* infection of migrating CD4⁺ T cells and of persistent immune activation would be removed. However, there could be problems with this approach. For example, it is unclear where the released virions would end up after being flushed out of the FDCs. It would certainly be an advantage if they could now be properly neutralized and ultimately destroyed. Many of these issues could be addressed using animal models.

In all, it is clear that lymphoid organs play a major role in the pathogenesis of HIV disease. Heath *et al.* have added another piece to the complex mosaic of host-virus interactions that are responsible for disease evolution. ▀

Lewis K. Schrager and Anthony S. Fauci are at the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA.

1. Heath, S. L., Tew, J. G., Tew, J. G., Szakal, A. K. & Burton, G. F. *Nature* **377**, 740–744 (1995).
2. Fauci, A. S. *Science* **262**, 617–622 (1993).
3. Pantaleo, G. *et al. Immun. Rev.* **140**, 105–130 (1994).
4. Pantaleo, G. & Fauci, A. S. *A. Rev. Immun.* **A 13**, 487–512 (1995).
5. Embretson, J. *et al. Nature* **362**, 359–362 (1993).
6. Pantaleo, G. *et al. Nature* **362**, 355–359 (1993).
7. Piatak, M. *et al. Science* **259**, 1749–1754 (1993).
8. Ho, D. D. *et al. Nature* **373**, 123–126 (1995).
9. Wei, X. *et al. Nature* **373**, 117–122 (1995).
10. Sabin, A. B. *Brit. J. exp. Path.* **16**, 169–188 (1935).
11. Phipps, R. P., Mandell, T. E. & Tew, J. G. *Cell. Immun.* **57**, 505–516 (1981).

PALAEONTOLOGY

Of it, not above it

Stephen Jay Gould

In 1867, Ralph Waldo Emerson wrote of the worm, who, “striving to be man/ Mounts through all the spires of form”. As the paper by Chen *et al.* on page 720¹ reminds us, the fossil record mocks our cultural expectations and psychological hopes for construing evolution as a steady rise in progress, with humans as a predictable apogee.

No phenomenon of life's history seems less suited to Emerson's model than the Cambrian Explosion, the remarkable episode which lasted only 10 million years (from 530 to 520 million years ago) and featured the first appearance in the fossil record of effectively all modern animal phyla, including annelid worms and chordates. Charles Darwin faced this challenge to his gradualistic preferences with characteristic honesty, writing in the first edition of the *Origin of Species*²: “The case at present must remain inexplicable; and may be truly urged as a valid argument against the views here entertained”. As usual, he attributed the apparent rapidity to imperfections of the fossil record and speculated that recognizable

ancestors of modern phyla must have inhabited older seas and not been preserved: “During these vast, yet quite unknown, periods of time, the world swarmed with living creatures”.

The earliest Cambrian is divided into three parts called, from oldest to youngest, Manakayan, Tommotian and Atdabanian, to honour Russian localities where early Cambrian rocks are particularly well exposed. The Manakayan contains many fossilized bits and pieces of cousins and precursors, but not many remains of major modern phyla. The Manakayan therefore pre-dates the Cambrian Explosion. By the end of the Atdabanian, virtually all modern phyla had made their appearance. The Cambrian Explosion therefore spans the Tommotian and Atdabanian.

Contrary to Darwin's expectation that new data would reveal gradualistic continuity with slow and steady expansion, all major discoveries of the past century have only heightened the massiveness and geological abruptness of this formative event for the kingdom Animalia. The soft-

bodied fauna of the Burgess Shale, discovered by C. D. Walcott³ in the Canadian Rockies in 1909, provided good insight into the full range of diversity achieved at this most auspicious beginning. Later interpretations by Whittington and students^{4–6} showed that Walcott had underestimated the true extent of diversity by classifying all Burgess fossils into extant groups, whereas many lie outside modern anatomical ranges. But the Burgess Shale dates from the subsequent Middle Cambrian, thus providing a bit of potential time for diversification after the main event of the explosion itself.

During the past decade, however, the discovery and development of another fauna of marvellously preserved soft-bodied Cambrian organisms at Chengjiang in China has proven that full diversity was reached within the explosion itself — for the Chengjiang fauna contains a range as broad as Burgess (including many of the same genera), but dates from older rocks of the Atdabanian — that is, within the time of the Cambrian Explosion.

For understandably parochial reasons, we are forever seeking criteria that might make our own ancestors special or inherently more progressive — and, case by case, the fossil record quashes our pretensions. Older textbooks proclaim that our phylum, the Chordata, did not appear until the subsequent Ordovician period, and that this later evolution must imply advanced status. But the Burgess Shale contains a chordate, the genus *Pikaia*, misidentified by Walcott as a polychaete annelid. However, *Pikaia* remains in limbo, for no comprehensive anatomical description has yet been published.

Chen and colleagues¹ discovery and description of a beautifully preserved and unambiguously identified chordate from the still earlier Chengjiang fauna now seals the fate of this misguided effort in asserting specialness for our ancestry. Chordates arose in the Cambrian Explosion. The only post-Cambrian appearance for a phylum belongs to the Ectoprocta, a group of marine colonial organisms prominent in the Palaeozoic fossil record, relatively inconspicuous today, and utterly unknown to the world at large (however beloved by all palaeontologists). Ectoprocts appear in the Ordovician period, and I will take refuge in Darwin's argument to predict that we just haven't found the Cambrian representatives yet.

The new Chengjiang chordate, *Yunnanozoon lividum*, described by a wonderfully international team of five authors from four maximally diverse and distant nations (invertebrate palaeontology has always been a remarkably ecumenical and cooperative enterprise), is so well preserved that its affinity within the Chordata can also be specified. Chordates are divid-

ed into three major lines — the tunicates, the cephalochordates (represented today by *Amphioxus* and its relatives), and the craniates (including all vertebrates). *Yunnanozoon*, with its metameric gonads and anteriorly extended notochord, belongs to the cephalochordates. As the authors note, the fact that one major division is already differentiated by unique characters within the Cambrian Explosion probably indicates that the other two divisions existed then as well — and that not only the phylum Chordata itself, but also all its major divisions, arose *within* the Cambrian Explosion. So much for chordate uniqueness marked by slightly later evolution.

Other discoveries continue to highlight the speed and magnitude of the Cambrian Explosion. Bowring and colleagues⁷ recently provided our first rigorous radiometric dates for the event — and 'fast' turns out to be much faster than anyone ever thought. The Tommotian and Atdabanian span only 6 to 10 million years, not up to 30 as previously argued. Meanwhile, new fossil discoveries have extended the range of several more phyla, including the Tardigrada⁸ and Pentastomida⁹, into the Cambrian. The case of the pentastomes is particularly remarkable, for these parasites of vertebrates had no previous fossil record at all, while their virtual exclusion to vertebrates as modern hosts made later evolution quite plausible.

The Cambrian Explosion occurred in a geological moment, and we have reason to think that all major anatomical designs may have made their evolutionary appearance at that time. Books have been written on the potential meaning of this remarkable phenomenology for revised views of evolution, ecology and development. Speculative and tendentious as much of this work may be (including my own), let us rejoice in the strangeness and elegant documentation of the phenomenology itself. Our own phylum, as *Yunnanozoon* proves, forms part of this universal story. As for our place in the history of life, we are of it, not above it. □

Stephen Jay Gould is in the Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA.

The radiation equation

Sasha Madronich

ONE of the consequences of decreasing levels of stratospheric ozone is that more ultraviolet radiation (UVR) will reach the Earth's surface, with possibly harmful consequences for both plant and animal life. The increases in UVR corresponding to the declines in stratospheric ozone at high and middle latitudes of the globe have been estimated¹. Yet it could be that such increases will have little effect on the biosphere if they are smaller than the natural fluctuations in UVR caused by other factors that affect atmospheric transmission, among which cloudiness is probably the most important.

On page 710 of this issue², Lubin and Jensen combine satellite-based measurements of ozone and cloudiness to estimate the UVR levels at the Earth's surface. They conclude that if reductions of stratospheric ozone continue at recently measured rates, many temperate and high-latitude locations will experience rises of monthly averaged UVR levels above cloud-induced variability within a few decades. For some locations, they find that this point has already been exceeded.

Lubin and Jensen chose as their threshold the point when ozone-related rises in UVR exceed the standard deviations of cloud-induced interannual UVR variability by a factor of 1.96. This is reasonable as a signal-to-noise threshold for potential effects on the biosphere. Local populations and ecosystems may be well adapted to cope with UVR fluctuations within a certain range defined by normal ozone levels and natural cloud variability; but when ozone reductions bring the mean levels of UVR above this natural range, the biological effects are likely to be considerable.

As the authors recognize, their study uses relatively short global records for ozone (about 13 years) and cloud cover (about 5 years), allowing only a preliminary glimpse of changes in the UVR environment. Also, the use of monthly averaged cloudiness tends to smooth out some of the shorter-scale variability, which would obviously be higher if computed, for example, on instantaneous, hourly or daily bases. Additional variability stems from the natural variability in atmospheric ozone content, and from the fact that the highest UVR levels are usually attained under relatively cloud-free skies. Although cumulative (for instance monthly) UVR exposure is generally thought to be important, the occurrence of unusually high peak values may be of added significance for biological effects that have non-linear dose-response relationships, or when there is a tendency for

exposure to occur under clear skies (as, for instance, in the human preference to take part in recreational activities on sunny days).

Nonetheless, Lubin and Jensen's results provide a useful measure of how reductions of ozone will affect the UVR environment. The geographical and seasonal distributions are largely consistent with the increases in UVR estimated for cloud-free skies, and reinforce the view that UVR increases will be most pronounced at the middle and high latitudes of both hemispheres. Particularly noteworthy is the indication that summertime levels of UVR may already be exceeding the background cloud variability in some parts of Europe, central Asia, North America and the Southern Hemisphere poleward of 30 degrees.

At other locations, primarily in the tropics where reported ozone trends are small, and in those regions of middle and high latitudes that normally experience highly variable cloudiness, the ozone-induced increases in UVR may be less meaningful. In particular, the rapid achievement of the threshold in some tropical regions should be understood to result mostly from low cloud variability rather than large reductions in ozone levels. Very small-scale geographical variations, at the 2.5×2.5 degree latitude-longitude pixel size of this study, should be viewed with some caution because of the relatively short data record.

The complex temporal and spatial distributions reported by Lubin and Jensen underline the need for continued and expanded monitoring of UVR at the Earth's surface, so as to characterize its climatologically normal averages and variabilities, and to identify long-term trends and their causes. The importance of UVR as an environmental variable is now well established, and studies continue to confirm its involvement in, among other things, induction of skin cancer, eye damage, immune-system suppression and plant damage, in both terrestrial and aquatic ecosystems³. Yet relatively little is known about geographical and seasonal distributions of UVR; and because of the absence of a reliable, long-term monitoring network, information has already been lost about the natural baseline against which

1. Albritton, D. & Watson, R. (eds) *Scientific Assessment of Stratospheric Ozone: 1994. WMO Global Ozone Research and Monitoring Project, Rep. 37* (World Meteorological Organization, Geneva, 1995).
2. Lubin, D. & Jensen, E. H. *Nature* **377**, 710–713 (1995).
3. van der Leun, J., Tevini, M. & Tang, X. (eds) *Environmental Effects of Stratospheric Ozone Depletion — 1994 Update* (United Nations Environmental Programme, Nairobi, 1994).

1. Chen, J.-Y., Dzik, J., Edgecombe, G. D., Ramsköld, L. & Zhou, G.-Q. *Nature* **377**, 720–722 (1995).
2. Darwin, C. R. *On the Origin of Species by Means of Natural Selection* (Murray, London, 1859).
3. Walcott, C. D. *Smithsonian Misc. Coll.* **57**, 1–16 (1910).
4. Whittington, H. B. *The Burgess Shale* (Yale Univ. Press, New Haven, 1985).
5. Gould, S. J. *Wonderful Life: The Burgess Shale and the Nature of History* (Norton, New York, 1989).
6. Briggs, D. E. G., Erwin, D. H. & Collier, F. J. *The Fossils of The Burgess Shale* (Smithsonian Institution Press, Washington DC, 1994).
7. Bowring, S. A. et al. *Science* **261**, 1293–1298 (1993).
8. Walossek, D., Müller, K. J. & Kristensen, R. M. *Sixth International Symposium on Tardigrada* (Cambridge, UK, 1994).
9. Walossek, D. & Müller, K. J. *Trans. R. Soc. Edinb., Earth Sci.* **85** (1994).

to evaluate current and future trends.

The net destruction of stratospheric ozone stems largely from anthropogenic emissions of chlorine-containing compounds. With full international compliance to the Montreal protocol and its subsequent amendments and adjustments, the amount of chlorine in the stratosphere is expected to peak near the turn of the century, then return gradually to pre-1980s levels over the next four or five decades. But the processes that control concentrations of stratospheric ozone are still not understood completely, particularly with respect to their interactions with

natural perturbations, such as volcanic eruptions, and human-driven effects. Perhaps the biggest uncertainty in Lubin and Jensen's study concerns the validity of extrapolating past ozone trends into the next century — an issue that has as much to do with humankind's ability to restrain its growing influence on atmospheric composition as it does with atmospheric processes. □

Sasha Madronich is in the Atmospheric Chemistry Division, National Center for Atmospheric Research, PO Box 3000, Boulder, Colorado 80307, USA.

NEUROBIOLOGY

The bee's needs

Rodney J. Douglas

ANIMALS must learn a substantial part of their behavioural repertoire through trial and error interactions with a dynamic and uncertain environment. A simple neural substrate for such a learning process is described in a paper by Montague and colleagues that appears elsewhere in this issue (*Nature* 377, 725–728; 1995). Their particular contribution is to show how bumblebee foraging can be interpreted in a number of contexts: macroscopic behaviour, reinforcement, artificial neural networks, and a combination of observed and plausible neuronal physiology.

Since the middle of this century psychologists have been trying to extract the fundamental principles of learning and memory that govern the acquisition of different behaviours. Their lead has been followed by neuroscientists, engineers and computer scientists, whose interests in the problem vary between biological explanation and practical application.

There is now general agreement that elementary associative processes are the building blocks of learning and that these associations can be established by reinforcement, in which the occurrence of a stimulus in the proper relation to a response increases the probability that the response will occur again in a similar situation. The learning of behaviour by reinforcement has at least two aspects. The appropriate internal sensory and motor representations of the world must be established, and then combined into the optimal sequences of behaviour that lead to the

goal. The reinforcement can be provided by an external teacher, but there is particular interest in cases where this supervision occurs internally, so that learning is autonomous.

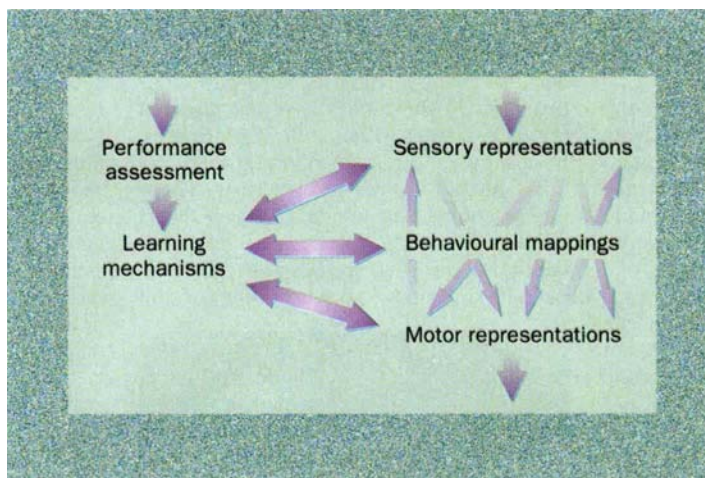
The model of Montague *et al.* is expressed as a computer simulation of the bee in an environment of flowers that

by reinforcement. The authors postulate that an identified neuron in the bee, VUMmx1, provides a prediction of reward to the bee. The predictor neuron combines actual nectar reward signals with its own prediction to change its evaluation of colour sensory input, thereby improving future predictions of reward. The authors show that the model bee's behavioural strategies in response to various nectar distributions across the flowers is in good agreement with observations of real bees. What computational principles underlie this success?

Most models of biological neuronal networks assimilate experience in the efficacy, or weights, of their synaptic connections. New experience is incorporated into the network by modification of the synaptic weights, and a popular learning rule for doing this is hebbian learning, which senses the correlation between the activities of pre- and postsynaptic neurons. This rule is unsupervised, because it does not require a teacher to assess the network's performance. These unsupervised, correlational methods learn statistical regularities in the sensory input. So they are useful in automatically extracting the most important features necessary to describe the input compactly. However, on its own, the simple hebbian method lacks two properties that are necessary to structure behaviour.

First, unsupervised methods are sensitive only to the degree of temporal correlation between inputs and not to their temporal order. But learning is asymmetric in time: sensory events that consistently precede reward stimuli come to act as predictors of the reward, whereas sensory events that follow the reward stimuli do not. This establishes the causal structure of the world; for example, landings on yellow flowers will be followed by access to nectar, which has a strong selectional advantage to the bee. Montague *et al.* have addressed this point by introducing a predictive modification of the hebbian rule into their model.

Second, the bee needs knowledge that is not inherent in the regularities of the input statistics (for example, the knowledge that yellow flowers are good things). These values are conveyed by a conceptually separate input signal, typically the assessment of an informed supervisor. In the bee model, a nectar reward neuron is an internal supervisor. It provides a reward signal that allows the predictive neuron to modify its synaptic weights so that the bee's strategy for



Autonomous learning in animals and robots. The animal, represented by the box, interacts with its dynamic and uncertain environment (the background) via perception and action. Successful behaviours require that the animal learn, or be granted, appropriate sensory and motor representations, which underlie perception and action. Learning mechanisms can organize the representations on the basis of regularities in the environment, but useful behaviours require knowledge of which actions are valuable. This information is provided by an internal supervisor that is able to assess the performance of the animal in the light of its needs.

contain more or less nectar, according to their colour. The bee has a cone of vision in which it is able to sense changes in percentages of colour. The bee needs nectar, and so computes a trajectory through the environment to flowers that will optimize its harvesting of nectar. The trajectory is a biased random walk based on current sensory input, and a prediction of nectar reward.

The bee's experience of reward is built

flower selection matches the expected reward.

The bee model is pleasing, but limited. The model bee has only one behaviour, the translation of the flower colour signal into a flight trajectory. The basic sensory and motor representations for achieving this task have been given to the system by the designers, rather than being learned. This approach is reasonable, but it does mean that the bee model optimizes only a single, simple strategy. Furthermore, although the neuronal aspects of the bee model are biologically plausible, the authors have made a number of assumptions that should now be verified experimentally.

Research in the field of autonomous learning has significant practical implications. The technology available for constructing autonomous robots is more advanced than our ability to have them express sophisticated behaviours. One reason for the disparity is that for some

years it was hoped that the behaviours of robots could be explicitly defined. That route has not been successful, in part because it assumed that the designer can have a complete understanding of the task that the robot must perform, and the environment in which it will operate. Such knowledge is rarely available, particularly for real environments.

The alternative approach is to engender behaviours implicitly. This method relies on the interplay between elements of behaviour available to the robot, and its environment. The hope is that these robots can be programmed by simple signals such as reward and punishment without specifying the task exactly. Montague *et al.* show how the analysis of simple biological systems can contribute to this goal. □

Rodney J. Douglas is at the Institute of Neuroinformatics, Gloriastrasse 32, Zürich 8006, Switzerland.

order of 0.07–0.7 mm wide)³. Some members of the heather family (Ericaceae), however, produce leafy shoots in which it is possible to detect annual growth, and among these is *Cassiope*⁴. At the end of each growing season the tightly appressed, drought-resistant leaves become smaller, only to expand to full size again at the height of the next growing season. The leaves persist from one season to another, so even over a period of 30–40 years there is a permanent record of annual growth in the form of periodic constrictions in the shoot profile. It is this feature that has been exploited by Havstrom *et al.*¹.

The retreat of ice at Ellesmere Island revealed an assortment of *Cassiope* shoots and only those that were clearly growing healthily when cut off in their prime by the ice development were selected for analysis. Twelve shoots were sufficiently well preserved to provide 25 years of growth record and a number of measurements were made on them, including leaf number, leaf length, total leaf biomass, stem length, and number of flowers or flower pedicels. All of these measurements could be expressed for each of the 25 years' growth. Because the plant still grows at this location, it was also possible to conduct similar morphological analyses on modern populations of *Cassiope* to provide a comparison with the fossil plants, which were dated by carbon-14 at between AD 1425 and 1665 (within the Little Ice Age period).

Mean values for flower number, leaf number and leaf biomass in the fossil shoots varied from year to year, but in a consistent manner. The twentieth year before the death of the fossil shoots, for example, gave low values for all three of these characters. The interannual variation, however, was lower in the fossil material than in modern counterparts measured back over 25 years, and the growth performance was generally poorer.

To translate growth measurements into climatic terms, the authors look to work on the response of modern *Cassiope* to experimental manipulation⁵, in which micro-greenhouses were used to increase July temperatures of field populations of the plants by 2–4 °C. These studies showed that High Arctic *Cassiope* are limited in their growth by July temperature and provided evidence of a simple linear relationship between growth and July temperature. On the basis of the regression equation, it can be calculated that the July temperature experienced by the fossil *Cassiope* was about 0.7 °C lower than that of the present day in this area.

The final demise of the Ellesmere Island plants, however, does not seem to have been caused by any gradual decline in temperature, for there are no clear trends in the performance curves over time. It is more likely to have resulted from a sudden arrival of additional snow-

PALAEOECOLOGY

Back to the Little Ice Age

Peter D. Moore

SUCH is our obsession with our own species that when the prehistoric remains of a man emerge from the ice as an Austrian (or Italian) glacier retreats, the world's press gathers; even the discovery of the corpse of a mammoth in Siberia attracts a great deal of attention. Plants, alas, are less newsworthy, but they can tell us as much if not more about the history of our environment.

Take, for instance, findings from Ellesmere Island in Canada, where the retreat of ice has revealed the fast-frozen remains of a tundra dwarf-shrub, *Cassiope tetragona*, investigations of which are beginning to provide information about tundra climate 500 years ago. In the latest issue of *Functional Ecology*¹, Havstrom and colleagues report their analyses of the fossil material in which they found reduced growth and flowering rates during the Little Ice Age in this arctic environment. The larger promise is that such studies might provide the basis for predicting the response of arctic plants to projected future changes in climate.

Some climate models predict that the high latitudes will experience exaggerated increases in temperature if global climatic warming continues, which raises questions concerning the possible response of arctic vegetation. There are two main approaches to such questions: either the current vegetation can be monitored and experimentally manipulated, or one can look to palaeoecological techniques for documenting past vegetation changes in response to climate change.

Much progress has been made in the first area, but not so much in the second because palaeoecologists face particular problems in the tundra². Among these is the lack of trees which, in lower latitudes, have supplied direct documentation of climatic history in their growth rings. There is, of course, an abundance of dwarf shrubs (including birches, willows and heathers), but such species generally produce very narrow growth rings (of the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

New celebrity — the arctic dwarf-shrub *Cassiope tetragona*. Fossil material bears witness to the tundra climate of 500 years ago.

Heather Angel

fall which failed to melt in the following summer and which gradually compressed into ice. Rapid freezing and incorporation into the ice accounts for the excellent preservation of the fossils.

The work opens up a fresh approach to climatic reconstruction in the Arctic, but there are still many difficulties to be faced. Such material is rare in the Arctic and is clearly associated with a time of rapid local (and perhaps regional) climate change. To provide a fuller account of such changes, samples would be needed from a range of times in the past. The dating is also unsatisfactory and provides at best a kind of floating chronology similar to that which has often been described for dendrochronological studies; again, a complete sequence of plants from the present day backwards would be needed in order to secure the chronology.

Then there is the factor of the genetic diversity and phenotypic plasticity of tundra plant populations⁶. The microclimatic diversity and the climatic sensitivity of the Arctic seem to have generated a wealth of ecotypes within species, making the physiological extrapolation from present populations to past ones unreliable. Finally, there is the assumption that the fossil plants were limited by, and responded to, changes in July temperature. Vegetative growth in populations of *Cassiope* growing at lower latitudes is limited by nutrient supply⁷ rather than summer temperature, so past growth responses could be complicated by other types of environmental change apart from climate.

So, do these data provide a basis for projection into the future of Arctic vegetation using climatic predictions based on current models? The answer is 'yes', but only as a first approximation. The complexities of growth limitation in plants such as *Cassiope* may mean that increasing atmospheric deposition of airborne nitrates⁷ will have a greater effect on their growth than raised temperature. Differential stimulation of potential competitors may also mean that any growth increase in low-growing dwarf shrubs will soon be counteracted by the shading effect of invasive pine and spruce. Predictive models of vegetation change cannot, unfortunately, be based on information about a single species. □

Peter D. Moore is in the Division of Life Sciences, King's College, Campden Hill Road, London W8 7AH, UK.

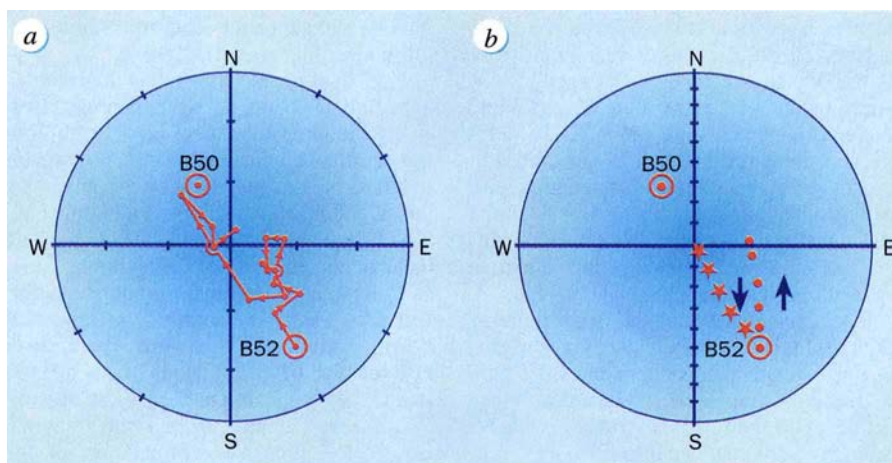
Storm in a lava flow?

Andy Jackson

In a paper early this summer, Coe *et al.*¹ provided yet more evidence to substantiate their intriguing and controversial observations of the past decade^{2,3} which suggest that the Earth's magnetic field can change extremely rapidly. These changes occur at rates (up to 6° of directional change per day) which are stupefying when contrasted with what we observe now and with what is seen in the records of the past four centuries (some fractions of a degree per year). The changes apparently occur when the magnetic field is in a perturbed state, in the process of undergoing a reversal from one polarity to another, and have occasionally been recorded in ancient lava flows at Steens Mountain, Oregon, a giant pile of lava

First, a little more background to the issue. Reversals of the magnetic field occur roughly every million years or so, and appear to take about 10,000 years to complete, a figure that ties in nicely with our estimate of the time that the field would take to decay through ohmic heating if the energy source that sustains it were suddenly removed. The fluctuations that occur during the 99 per cent of the time that the field has a stable polarity are small.

But Coe and colleagues found that the magnetic field locked up when the 16.2-million-year-old lava flows at Steens Mountain cooled through their Curie point, indicated changes of up to 6° per day, a factor of 1,000 larger than anything



Measurements and predictions of magnetization directions for the first jump in magnetization recorded in the lava flows at Steens Mountain, plotted on a stereographic projection. *a*, The measured directions move from the pre-jump stable direction (B52) to the post-jump stable direction (B50). *b*, Calculated directions caused by a magnetic storm, as in the new paper by Ultré-Guérard and Achache⁴, gradually alter the directions away from the B52 direction. Dots indicate a preferential orientation; stars are for the orientation determined by the pre-jump dipole axis.

almost 1 km in thickness. This site was where Coe and colleagues made their observations.

Writing in *Earth and Planetary Science Letters*, Ultré-Guérard and Achache⁴ have now joined the ranks of the sceptics who find these observations difficult to reconcile with an origin in the Earth's fluid core. They propose instead that the rapid fluctuations originate from an ancient magnetic storm, much like the ones that hit the Earth nowadays and occasionally cause communications blackouts and disruptions to electrical supplies. Their suggestion, which will no doubt cause more controversy within the community, adds yet another chapter to a complex history which, once unravelled, could profoundly affect our understanding of the geodynamo. As such, it deserves careful scrutiny.

observed today. Their inferences were based on the observation of a coherent and systematic change in the direction of magnetization within one flow as samples were taken closer to its centre. As the centre of the flow cooled last (roughly 10 days after the edges⁵), it records the later magnetic field direction. The sheer enormity of the difference between the rates of change implied in this work and those normally encountered led to a fair degree of incredulity within the scientific community.

From the beginning, there were some who doubted the veracity of the observations, arguing that the signals locked up in the rock's magnetization do not reflect true changes in the Earth's field (for example, chemical alteration or overprinting by a subsequent lava flow can ruin the recordings). But there are many reasons

- Havstrom, M., Callaghan, T. V., Jonasson, S. & Svoboda, J. *Funct. Ecol.* **9**, 650–654 (1995).
- Ritchie, J. C. *Past and Present Vegetation of the Far Northwest of Canada* (Univ. Toronto Press, 1984).
- Warren-Wilson, J. *Ann. Bot.* **28**, 72–76 (1964).
- Callaghan, T. V., Carlsson, B. A. & Tyler, N. J. C. *J. Ecol.* **77**, 823–837 (1989).
- Havstrom, M., Callaghan, T. V. & Jonasson, S. *Oikos* **66**, 389–402 (1993).
- Crawford, R. M. M. & Abbott, R. J. *Bot. Acta* **107**, 271–278 (1994).
- Moore, P. D. *Nature* **374**, 117–118 (1995).

why one might take the observations at face value, not least because of the very thorough job done by Coe *et al.* in discounting many possible types of artefact.

The main problem lies with assigning the origin of the magnetic changes to the Earth's core. First, given that magnetic signals must propagate the 2,900 km from the core to the Earth's surface through the electrically conducting mantle, could such a rapid signal be observed at Steens Mountain, given that the mantle's conductivity has the effect of smearing out even the sharpest impulsive change? Current experimental results for mantle conductivity indicate that field changes traversing the mantle which are dipolar (equivalent to flipping a bar magnet in the core) cannot generate changes more rapid than 2° per day. Only more complex magnetic fields with smaller lengthscales can be responsible for changes faster than this.

Second, a core origin causes problems with energetics: fast changes in the magnetic field require either rapid velocities at the top of the core (with large kinetic energies), or slower velocities and extremely small-scale magnetic fields (large time variations can be generated by moderate flows acting on magnetic fields with large spatial gradients). But even the latter picture is unattractive energetically, because small-scale fields generate a great deal of ohmic dissipation.

There is a way out of this dilemma. One thing that is universally agreed is that during a reversal the intensity of the field is greatly diminished, sometimes by an order of magnitude. So rather than an intensity of typically 50 microteslas, during the reversal at Steens Mountain it measured only 6 microteslas. At present the dipole orientation is responsible for fixing the orientation of the ring current in the magnetosphere, which usually lies at about 12 Earth radii and which is ultimately the source of magnetic storms when changes in the solar wind affect the Earth.

The effect of a weakened internal magnetic field, such as that recorded at Steens, is twofold. It causes the magnetosphere to contract (to about half its size) and consequently deliver stronger external magnetic fields to the Earth's surface, by about a factor of two. More pertinently, these external magnetic fields can have much more of an effect on the local field direction when a storm occurs, because they are of comparable size to the internal field, rather than being mere perturbations⁶.

The figure on the previous page shows one of the Steens lava flows representing a time when fast changes have been observed. The magnetization direction moves rapidly through about 80° from a stable pre-jump direction to a different stable post-jump direction. But compare

this with a prediction⁴ for what happens to the magnetization direction when a typical magnetic storm is in progress. This can be calculated in two ways — for the case when the ring current's orientation is fixed by the pre-jump palaeomagnetic dipole axis, or for an optimized orientation. Because of the great controversy regarding the morphology of the magnetic field during a reversal (whether it remains dipolar or has a more complicated geometry), either calculation is tenable. Either way, the feasibility of the proposed mechanism as an explanation for the observations is clear.

Coe *et al.* did not neglect the possibility of a magnetic storm. But they dismissed the idea because they envisaged the external field being responsible for the entire 80° jump from pre-jump to post-jump stable directions. Ultré-Guérard and Achache's suggestion is that the post-jump direction is coincidentally in the same direction as that produced by a magnetic storm superimposed on a stable pre-jump internal magnetic field.

The proposal is not far-fetched, although it has its shortcomings. Certainly the magnitudes are about right, but the duration required for such storms (6–10 days) is rather long by modern standards. What might seem rather improbable is the idea that two lava flows should have occurred at such unfortunate times as to have captured two magnetic storms. But set against the many lava flows that do not show signs of storms, the ratio is not too far off what would be expected from the modern frequency of storms. What does perhaps stretch the imagination is the supposed coincidence of the storm directions and the post-jump magnetization directions, which are physically unrelated.

Models of the Earth's internal dynamo, aided by increasing computing power, have now progressed to the stage of being able to generate a realistic simulation of the Earth's field for the equivalent of 40,000 years (and the latest model⁷ actually exhibits a reversal). As work continues, the unravelling of palaeomagnetic data such as those from Steens Mountain will, I suggest, be crucial to the interplay between observation and theory. □

Andy Jackson is in the Department of Earth Sciences, University of Leeds, Leeds LS2 9JT, UK.

1. Coe, R. S., Prévot, M. & Camps, P. *Nature* **374**, 687–692 (1995).
2. Prévot, M., Mankinen, E. A., Grommé, C. S. & Coe, R. S. *Nature* **316**, 230–234 (1985).
3. Coe, R. S. & Prévot, M. *Earth planet. Sci. Lett.* **92**, 292–298 (1989).
4. Ultré-Guérard, P. & Achache, J. *Earth planet. Sci. Lett.* **135**, 91–99 (1995).
5. Camps, P., Prévot, M. & Coe, R. S. *C. R. Acad. Sci. Paris* **320**, 801–807 (1995).
6. Starchenko, S. V. & Shcherbakov, V. P. *Dokl. Acad. Nauk SSSR* **321**, 69–74 (1991).
7. Glatzmaier, U. & Roberts, P. H. *Nature* **377**, 203–209 (1995).

Away with rain

"INTO each life some rain must fall", but Daedalus wants to dodge his share. His new form of rain-proofing does not intercept the falling rain drops: it deflects them. It is based on the fact that the vapour pressure of water rises rapidly with temperature. If you could make one side of a rain drop hotter than the other side, the difference in vapour pressure across it could amount to a significant fraction of an atmosphere. This unbalanced pressure would accelerate the light droplet sideways.

The DREADCO rain deflector therefore shines a powerful beam of infrared radiation up into the falling rain. The beam has a carefully designed axial intensity gradient, strongest in the centre, but declining steadily towards the outside. A rain drop falling into the beam is thus preferentially heated on the side nearest the beam centre, and is deflected sideways out of the beam.

Cunningly, the rain deflector employs a pulsed infrared laser, whose intense pulses are much shorter than the time-constant for thermal conduction within a drop. Each pulse thus imposes a strong thermal gradient across its target drops, established too quickly to be evened out by internal heat transfer.

One problem is that the rain can come from almost any angle, depending on the strength and direction of the wind. So DREADCO's pilot model has its laser on a steerable mounting. An imaging sensor detects the infrared backscatter from the illuminated drops, and continuously orients the laser to maintain an apparent circular distribution. This keeps the beam pointing into the approaching rain.

The rain deflector will be too complex and power-hungry to displace the personal umbrella. But in fixed installations it should transform the outdoor life, particularly in Britain. Pavement cafés, cheerful clothing, dry outdoor events, predictable sporting fixtures, open-topped cars, all will at last become practical. A vehicle-mounted deflector could also be used to deflect fog, which is made of smaller water droplets. The pulsed infrared beams of DREADCO's anti-fog lamp will elbow aside the swirling mists, giving clear vision in the trickiest driving conditions. In the same way, an upward-gazing, long-range, low-divergence model might even be able to deflect the clouds. The lucky user could bore a hole through an overcast sky in the direction of the Sun, and illuminate himself in a bright beam of personal sunlight. Unscrupulous evangelists will rush to acquire the apparatus, so as to demonstrate their apparent high favour with the Almighty.

David Jones

Condensed-phase nanotubes

SIR — Carbon nanotubes¹ are conventionally prepared by gas-phase deposition in arc reactors, or by thermal decomposition of hydrocarbons². We have found that the synthesis of nanotubes from a condensed phase is also possible. Specifically, we have prepared multi-walled nanotubes by passing an electric current between carbon

electrodes in molten lithium chloride.

A simple carbon crucible, made by drilling a cylindrical hole (2.5 cm × 3 cm) in the centre of one face of a 5 × 5 × 5-cm cube of high-purity carbon, was filled with around 1 g of lithium chloride and heated to its melting point (600 °C) in air. A 3-mm-diameter high-purity carbon rod was immersed in the melt (to a depth of about 1.5 cm) in order to form the cathode of an electrolysis cell; the crucible acts as the anode. Around 30 A was maintained through the melt for 1 min. The immersed surface of the cathode rod was eroded during electrolysis, small pits appeared and particulate material (10–30 mg) was found dispersed throughout the melt. After cooling, the solid melt was added to water in order to dissolve the lithium chloride and react with the residual lithium metal. The mixture was set aside for 4 h, then toluene was added to the aqueous suspension and the whole was agitated for several minutes.

After allowing the mixture to settle, it was found that all the solid material seemed to have passed into the toluene layer. The aqueous and organic layers were separated by careful decanting. After sonication, droplets were removed from the toluene fraction and deposited on a carbon grid for transmission electron microscopy (TEM). A microscopic solid remained after removing the toluene by evaporation.

Low-resolution TEM studies reveal the particulate material to consist of multi-walled nanotubes and spheroidal carbon particles of 30–50 nm diameter; the latter tend to cluster. These are onion-like polyhedral particles, similar to those observed by Iijima³ and Ugarte⁴. Some of the nanotubes (2–10 nm in diameter) are very long (>500 nm) and show evidence of encapsulated material (lithium chloride, oxide or possibly lithium metal). The nano-

tubes consist of 5–20 concentric layers and, unlike those produced by the vapour arc/plasma technique^{1,2}, are not always straight, but tend to form loops. The tubes collapse on prolonged irradiation (around 15 min) in the electron beam, forming irregular distorted tube-like structures. We believe that the formation of nanotubes during electrolysis may have important implications for continuous methods of nanotube production, as well as, for example, facilitating encapsulation of material within the nanotubes.

W. K. Hsu

J. P. Hare

M. Terrones

H. W. Kroto

D. R. M. Walton

School of Chemistry and Molecular Sciences,

University of Sussex, Brighton BN1 9QJ, UK

P. J. F. Harris

Chemistry Department, University of Reading, Reading RG6 2AD, UK

1. Iijima, S. *Nature* **354**, 56–58 (1991).

2. Ebbesen, T. W. & Ajayan, P. M. *Nature* **358**, 220–222 (1992).

3. Iijima, S. *J. Cryst. Growth* **5**, 675–683 (1980).

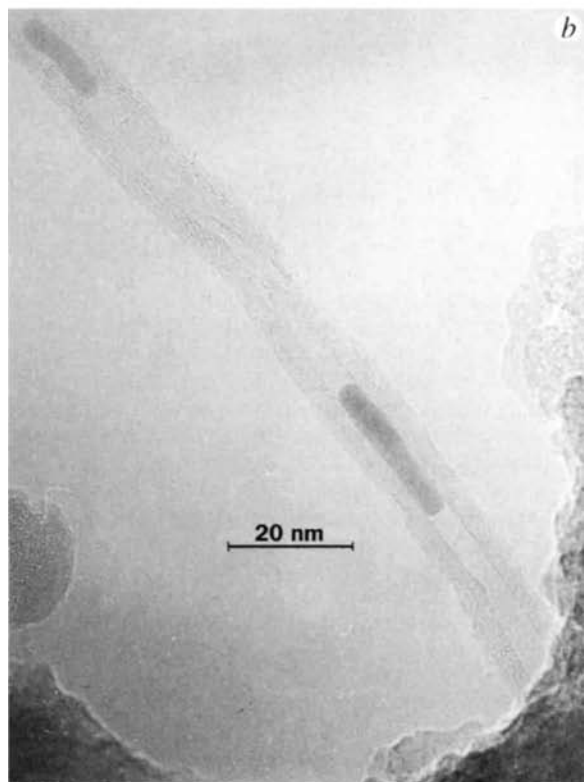
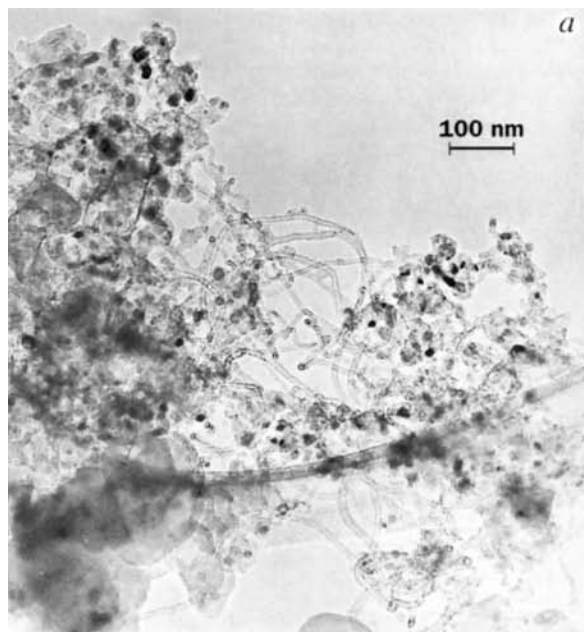
4. Ugarte, D. *Chem. Phys. Lett.* **207**, 473–486 (1993).

Evaporation losing its strength

SIR — Over the past several decades, the observed increase in mean temperatures over much of the world has been a result of a disproportionate increase in nighttime temperatures¹. Here we show that, in the United States and the former Soviet Union, the evaporation of water, as measured by pan evaporimeters — simple devices consisting of a pan of water, a device to measure the water needed to return the surface to a predetermined level, and a rain gauge — has decreased. These decreases seem to be closely related to the same factors that are causing the trends in temperature.

Pan evaporation is a simple measurement of complex meteorological interactions. Although pan evaporation cannot fully represent lake evaporation and is even less indicative of ground evapotranspiration, analysis of observed trends in pan evaporation can provide considerable insights into current climate change and the impact this change may be having on agriculture and water resources.

A wide variety of evaporation pans have been used in both Russia and the United States since the nineteenth century^{2,3}. Because pan type can affect the measurement, assessment of the homogeneity of the data is important before any trend analysis is carried out.



a, Low-resolution TEM image of a sample of the particulate material collected from the LiCl melt after electrolysis. The curly structure is typical. b, High-resolution TEM image revealing encapsulated material within one of the nanotubes.

For the former Soviet Union, analysis of station history documentation³ yielded a network of 190 homogeneous evaporation-pan reporting stations, with data starting in 1951. The digital US pan-evaporation network was determined to be homogeneous by the good agreement in trends between all 746 stations and a small subset whose station history indicated no possible inhomogeneity-inducing changes.

The data used are seasonal averages of mean daily pan evaporation from May to

September for the United States, and months when there was no ice in the pans for the former Soviet Union. Pan evaporation for four of five geographical regions show significant downward trends over the past 50 years (see figure). Area-averaged time series of pan evaporation for each of these regions were compared with similar area-averaged time series for various climatic parameters, including mean minimum temperature, mean maximum temperature, diurnal temperature range (DTR), total cloud cover, and precipitation. Of these, pan evaporation was most highly correlated with DTR, with a mean r^2 of 0.48. Widespread decreases in DTR over the past several decades are also among the primary findings of analyses of asymmetrical trends in daily maximum and minimum temperature¹. Pan evaporation and DTR have varied similarly in recent decades.

The downward trend in pan evaporation over most of the United States and former Soviet Union implies that, for large regions of the globe, the terrestrial evaporation component of the hydrological cycle has been decreasing. This partially explains increases in runoff over the past two decades in the European part of the former Soviet Union⁴ and the northern United States⁵, and corresponds well with both decreases in maximum summer temperatures over these regions¹ and a decrease in growing-season degree-days (annual sum of daily temperatures above

5 °C) over the Siberian and European former Soviet Union⁶.

Karl *et al.*¹ speculated that increases in cloud cover, especially low cloud cover, may explain the decreases in DTR. Our results support this speculation, as pan evaporation has been decreasing and is correlated negatively with cloud cover with a mean r^2 of 0.34 for the five regions. Aspects of the temperature and hydrological cycles have changed in tandem over the past 50 years, and the features of this recent climate change include decreases in potential evaporation.

T. C. Peterson

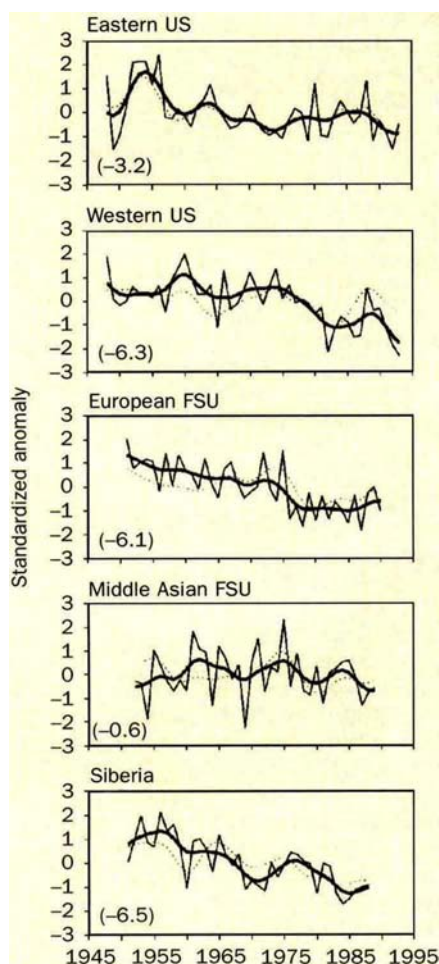
National Climatic Data Center/NOAA,
Asheville, North Carolina 28801, USA

V. S. Golubev

State Hydrological Institute,
Sankt-Peterburg 100053, Russia

P. Ya. Groisman

Department of Geosciences,
University of Massachusetts, Amherst,
Massachusetts 01003, USA



Area-average pan evaporation (solid lines) and diurnal temperature range (DTR, dotted line, updated from Karl *et al.*¹) for three regions in the former Soviet Union and two regions in the United States. All time series are presented as standard deviation anomalies from the long-term mean values. The smooth curves result from 11-point binomial smoothing (for DTR only smoothed lines are presented). Linear trend estimates for these regions (in standardized anomalies per 100 yr) are shown in parentheses. They are significant at the 99% level except for the former Soviet Union Middle Asian region. The largest actual change in pan evaporation is in the western United States, where the area-averaged linear regression slope corresponds to a decrease in pan evaporation of 97 mm per warm season (May–September) during the past 45 years in a region with a mean pan evaporation of 1,130 mm per warm season.

Baboon fertility and social status

SIR — Packer and colleagues¹ reported that for the baboon population of Gombe, as for many other primate populations^{2–5}, high dominance status had salutary effects on several components of female reproductive success. They also reported that high status entailed certain reproductive costs, and proposed behavioural and endocrine mechanisms for their production. Here we address the generality of the costs reported, a potential methodological problem in the results, and the validity of the behavioural and endocrine assumptions of the proposed model of reproductive costs.

The conclusion that high rank is associated with increased risk of lifetime infertility¹ derived from the finding that, of Gombe females that had no successful pregnancies or matured extremely late, one was high-ranking, three had high-ranking mothers and the fifth was a 'social climber' who rose above her family to achieve high rank. The significance of the positive relationship between dominance and lifetime reproductive success was $P = 0.0675$ when all low-fertility females were included, and $P = 0.0354$ when the 2 of 37 females that never gave birth to live young were excluded from the data

set¹. Analyses for other sites^{3–6} have not reported cases of lifetime infertility.

Our 25-year longitudinal studies in Amboseli, Kenya, include 133 adult females in two wild-foraging groups (stable in size, 50% survival through two years) and a third that is partially food-enhanced (expanding in size, 90% survival). We have had only one case of lifetime infertility; this individual was not high-ranking. Similarly, neither of our two 'social climbers' was infertile. If individuals with non-lethal pathologies are more likely to survive when infant mortality is low, higher incidences of reproductive pathologies might be expected when infant survival is high. However, none of the 31 females in the expanding group has been infertile. Present evidence is, therefore, inadequate to determine either whether incidences of lifetime infertility or its relationship to social status differ significantly among groups or populations.

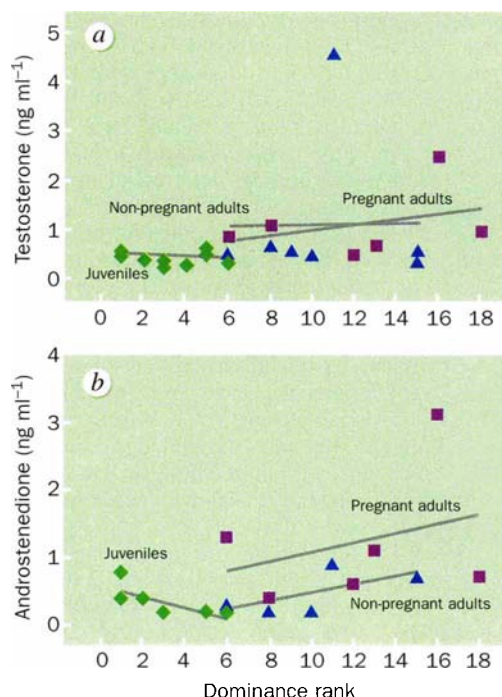
The most surprising observation made by Packer *et al.* was an association of high rank with high rates of spontaneous abortion¹, in contrast to a large literature that has implicitly or explicitly linked reproductive impairment and spontaneous abortion to the physical and

social stressors of social subordination⁷. Pregnancy was judged by the female's perineum turning pink¹, which is variable in timing and does not usually occur until 4–6 weeks into the 25-week gestation, precluding detection of early pregnancies and abortions. Therefore, while high status was clearly associated with mid- and late-pregnancy losses, the picture for early losses, and so total losses, cannot be discerned. Yet the bulk of pregnancy losses occur during this initial period and are those that seem to be most sensitive to stressors^{8,9}.

If the overall rate of stress-induced pregnancy loss was higher in subordinate females (as expected from the existing literature), the authors' apparently opposite finding might have resulted from their detection method, which is nonrandom with respect to pregnancy stage. Nor does the absence of a rank-related statistical difference in cycling time¹ remove this possibility, as the interval difference resulting from these early losses would be small. Wasser⁶ reported the same association between high status and high rate of spontaneous abortion at Mikumi; however, this same pregnancy detection method was used (personal communication).

For 474 conceptions in Amboseli in which we identified pregnancies just over two weeks from conception (less than a week after expected implantation), no relationship was found between dominance rank and rate of spontaneous abortion ($P>0.40$); in each group, higher-ranking females tended to have lower rates of fetal loss, not higher, as reported for Gombe and Mikumi.

Packer and colleagues speculated that high rank is associated with higher rates of agonistic encounters and, as a result, with high levels of circulating androgens which would produce the reported reproductive dysfunction of dominant adults¹. Although high rates of aggression might play a role in the initial establishment of dominance¹⁰, they are not normally associated with the maintenance of the stable dominance relationships^{10–12} characteristic of adult female baboons. Nor does high rank seem to be associated with high rates of agonistic inter-



The relationship between social dominance rank and serum androgen levels in female baboons (a, testosterone; b, androstenedione; diamonds, juveniles; triangles, non-pregnant adults; squares, pregnant adults). Dominance status is presented solely as status within age-sex class; an integrated female ranking would place juveniles scattered throughout the adult female ranking and reduce any slope and significance of status. Serum was collected and analysed as reported^{15,16}, with control for season, time of day and time from darting to blood drawing. Juveniles weighing less than 5 kg, females beyond the first half of pregnancy, and lactating females with young infants were not immobilized; the data, therefore, do not include individuals in these categories.

actions^{11,12}. Moreover, the literature does not support the notion that, among primates, increased levels of aggression within the normative range of social behaviour can raise androgen levels into a pathophysiological range¹³. Should such pathologically elevated androgen levels occur, all aspects of reproduction would be impaired¹⁴, whereas Packer and colleagues noted salutary effects of dominance on most measures of reproduction.

Packer and colleagues predicted that, over the entire rank range, higher

rank should be associated with higher androgen levels. No information exists regarding wild female primates. The closest approximations are our own limited data (see figure), in which no relationship of testosterone with rank is observed, while for androstenedione the pattern for juveniles supports this prediction ($P=0.047$), but that for adults runs counter to it ($P=0.17$). Thus, overall, the data give little support for the notion of dominant females being exposed to even slightly higher androgen levels than subordinates.

In conclusion, without dismissing the Packer *et al.* data or suggesting that high status is always advantageous, we note that dominance rank has almost always been found either to confer advantage or to be neutral with respect to a variety of fitness components, and the Packer *et al.* report provides some of the strongest evidence thus far for those advantages. Whether the higher miscarriage rates^{1,6} result from a methodological artefact or represent a difference among sites is unknown. In either case, neither the behavioural nor the endocrine hypotheses proposed to explain the apparent Gombe result are supported by available evidence, nor are the endocrine hypotheses proposed for Mikumi⁶ applicable to Gombe. Nevertheless, Packer and colleagues rightly insist that

any fitness relationship to dominance requires adequate physiological and behavioural mechanisms to arise and be maintained, a point often ignored in the evolutionary literature and a subject of much needed research.

Jeanne Altmann*

Department of Ecology & Evolution,
University of Chicago,
Chicago, Illinois 60637, USA

Robert Sapolsky*

Department of Biological Sciences,
Stanford University,
Stanford, California 94305, USA

Paul Licht

Department of Integrative Biology,
University of California,
Berkeley, California 94720, USA

*Also at: Department of Conservation Biology, Chicago Zoological Society, Brookfield, Illinois 60513, USA (J. A.); Institute of Primate Research, National Museums of Kenya, PO Box 24481, Karen, Nairobi, Kenya (J. A., R. S.).

- Packer, C. *et al.* *Nature* **373**, 60–63 (1995).
- Silk, J. in *Primate Societies* (eds Smuts, B. B. *et al.*) 318–329 (Univ. Chicago, 1987).
- Lee, P. C. *J. Zool.* **213**, 409–422 (1987).
- Altmann, J., Hausfater, G. & Altmann, S. A. in *Reproductive Success* (ed. Clutton-Brock, T. H.) 403–418 (Univ. Chicago, 1988).
- Smuts, B. B. & Nicolson, N. *Am. J. Primatol.* **19**, 229–246 (1989).
- Wasser, S. *Nature* **376**, 219–220 (1995).
- Wasser, S. & Barash, D. *Q. Rev. Biol.* **58**, 513–538 (1983).
- Myers, R. in *Psychoneuroendocrinology in Reproduction* (eds Zlotchella, L. & Panfili, P.) 555–574 (Elsevier, New York, 1979).

- Huisjes, H. *Spontaneous Abortion* (Churchill Livingstone, New York, 1984).
- de Waal, F. *Q. Rev. Biol.* **61**, 459–479 (1986).
- Silk, J. *Am. J. Primatol.* **12**, 479–495 (1987).
- Muruthi, P. thesis, Univ. Nairobi (1989).
- Monaghan, E. & Glickman, S. in *Behavioral Endocrinology* (eds Becker, J., Breedlove, S. & Crews, D.) 201–285 (MIT Press, Cambridge, MA, 1992).
- Ross, G. T. in *Williams Textbook of Endocrinology* 7th edn (eds Wilson, J. D. & Foster, D. W.) 206–258 (W. B. Saunders, Philadelphia, 1985).
- Sapolsky, R. M. & Altmann, J. *Biol. Psychiat.* **30**, 1008–1016 (1991).
- Glickman, S. E., Frank, L. G., Pavgi, S. & Licht, P. *J. Reprod. Fert.* **95**, 451–462 (1992).
- Glickman, S. E., Frank, L. G., Holekamp, K. E., Smale, L. & Licht, P. in *Perspectives in Ethology* Vol. 10 (eds Bateson, P. P. G., Klopfer, P. H. & Thompson, N. S.) 87–117 (Plenum, New York, 1993).
- Frank, L. G., Weldele, M. L. & Glickman, S. E. *Nature* **377**, 584–585 (1995).

PACKER REPLIES — Altmann *et al.* imply that reproductive performance is always positively correlated with dominance rank, when neither of their principal references^{2,4} provide compelling evidence of an overall relationship between rank and lifetime reproductive success. Indeed, Silk² concludes her review with the statement, “The lack of consistency among these results has created consider-

able controversy about the significance between dominance rank, reproductive success, and genetic fitness among females."

We emphasized that the presence or absence of just one or two high-ranking females with reproductive pathology would influence perception of the significance of rank. We reported on five such females at Gombe and cited examples in other populations and species. While it is interesting that such rank-related pathologies have not yet been observed at Amboseli, there is no reason to dismiss them from our data set. A more productive approach would be to ask why such females might be more common at one site rather than another. Perhaps the historically high rate of infant mortality at Amboseli⁴ has removed infertile animals from this population.

Altmann *et al.* suggest that, since field data cannot detect pregnancy for the first few days after conception, we may have missed many spontaneous abortions in the initial stages of pregnancy, and they predict that early abortions would be highest in low-ranking females. This prediction, however, is not confirmed by the Gombe data. If low-ranking females suffered a higher rate of undetected abortions immediately after implantation, they should take significantly more cycles to become obviously pregnant. However, once Gombe females resume cycling, there is no significant effect of rank on their interval to the next detectable pregnancy. Thus, early abortion is independent of rank at Gombe, and our miscarriage data are unbiased.

Further, we did not solely detect miscarriage on the basis of perineal coloration. Our analysis also included six miscarriages detected on the basis of delayed menstruation, and these all occurred in the first four weeks after conception (out of 50 miscarriages where the day of conception was known). There was no relationship between a female's rank and the gestation length of her failed pregnancy ($r=0.52$, $P=0.6081$, $n=48$ females of known rank). Our conclusion therefore remains unaltered: high-ranking Gombe females suffer more miscarriages, and this demographic pattern has been confirmed by Wasser⁶ in a third baboon population.

We suggested that high-ranking females might show higher levels of androgens, as these have been implicated in the reproductive pathologies of several other mammalian females. The data presented by Altmann *et al.* are too limited to provide a rigorous test of this hypothesis. Most miscarriages at Gombe were suffered by females ranking fourth or higher, and Altmann *et al.* provide no data for any adult female ranked higher than sixth. Further, their data on androstenedione levels in juvenile females actually

support our hypothesis. The regression of the log-transformed androstenedione levels on juvenile dominance rank is significantly positive ($P=0.0470$, $n=6$). Female dominance rank is established during the juvenile period and early exposure to androstenedione is believed to influence aggressiveness of adult females⁷. However, androgens are not the only hormones involved in aggressive behaviour, and competition may also be balanced by other physiological costs (see ref. 6). Further, reproductive pathologies at Gombe appear to be hit or miss—some high-ranking females have large numbers of surviving offspring whereas others have none. The most compelling data on reproductive failure would be from those specific adult females with some form of pathology.

The pattern in our miscarriage data does not result from our methodology. High-ranking females at Gombe and Mikumi⁶ suffer significantly higher rates of miscarriage during detectable pregnancies. At Gombe, the greater incidence of reproductive pathology and

the high miscarriage rate in high-ranking females essentially cancel the striking advantages of high rank from shorter post-partum amenorrhoea, higher infant survival and an earlier age of sexual maturation. These trends may or may not hold at all field sites, and contrasting results may prove useful in clarifying the costs and benefits of aggressive competition in different ecological settings. Although limited, the Altmann *et al.* data support our suggestion that female rank is correlated with androgen levels, androstenedione being highest in juvenile females of highest dominance rank. Wasser's⁶ study of yellow baboons and findings by Frank *et al.*¹⁸ on spotted hyaenas further emphasize the value of investigating the reproductive costs of aggressive competition in female mammals.

Craig Packer

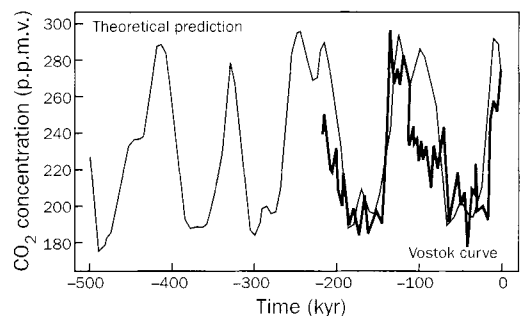
Department of Ecology, Evolution and Behaviour,
University of Minnesota,
1987 Upper Buford Circle,
Saint Paul, Minnesota 55108, USA

Predicting the Vostok CO₂ curve

SIR—To provide more than an *ad hoc* explanation of observed phenomena, a theoretical model must predict some as yet unobserved variability that places the model 'at risk' in some verifiable respect.

In our phenomenological dynamical-system model governing changes in global ice volume, atmospheric CO₂ and the deep ocean state over the late Cenozoic^{1,2}, as influenced by Earth-orbital and slow tectonic CO₂ forcing, free parameters were assigned to account for the complex variations in global ice volume revealed by $\delta^{18}\text{O}$ proxy evidence. This includes internal bifurcations to the 'ice age' around 2.5 Myr ago and to the main near-100-kyr-period variations around 0.7 Myr ago. As a side consequence, the main features of the variations in CO₂ over the past 218 kyr were deduced and these compare favourably with those determined by the Vostok core trapped-air measurements¹⁻⁴.

As the drilling of the Vostok core is now projected to approach a depth corresponding to roughly 500 kyr ago³, we now offer, as a further prediction of our model, the expected variation in CO₂ for



Variations over the Late Pleistocene of atmospheric CO₂ as inferred from the Vostok ice core³ (heavy line) and from the dynamical theory^{1,2,4} (thin line).

the additional 300 kyr (see figure). According to the theory, the character of the variations should change somewhat, showing a shorter-period fluctuation, with new minima near 300 and 370 kyr ago. At present, this model is the only one from which such a prediction can be made as a response to purely external forcing, dealing with CO₂ as a free, internal, dynamic variable^{5,6} rather than as a prescribed function (as in refs 7-9).

Barry Saltzman

Mikhail Verbitsky

Department of Geology &

Geophysics,

Yale University, PO Box 208109,

New Haven, Connecticut 06520-8109, USA

1. Saltzman, B. & Verbitsky, M. *Climate Dynamics* **9**, 1-15 (1993).
2. Saltzman, B. & Verbitsky, M. *Paleoceanography* **9**, 767-779 (1994).
3. Jouzel, J. *et al.* *Nature* **364**, 407-412 (1993).
4. Saltzman, B. & Verbitsky, M. *Nature* **367**, 419 (1994).
5. Saltzman, B. *Climate Dynamics* **1**, 77-85 (1987).
6. Saltzman, B. & Maasch, K. *Globi biogeochem. Cycles* **2**, 177-185 (1988).
7. Berger, A. *et al.* in *Long-Term Climatic Variations* (eds Duplessy, J.-C. & Spiridakis, M.-T.) 411-452 (Springer, Berlin, 1994).
8. DeBlonde, G. & Peltier, W. R. *J. Climate* **6**, 709-727 (1993).
9. Schlesinger, M. & Verbitsky, M. *Palaeoclimates: Data and Modelling* (in the press).

Birth of the clinical trial

Roy Porter

Quantification and the Quest for Medical Certainty. By J. Rosser Matthews. Princeton University Press: 1995. Pp. 195 \$39.50, £32.

Is medicine an art or a science? This was long a scholastic debate, with the weight of authority tending to come down in favour of the former — had not the very first of Hippocrates' aphorisms begun *ars longa vita brevis*? The issue was opened up again during the seventeenth-century scientific revolution, when progressive physicians attempted to develop iatro-mathematics. Yet such numerical medicine made but halting progress in a medical world dominated by clinicians whose prestige hinged on 'clinical judgement' grounded in vast but essentially personal wisdom.

As Dr Matthews recognizes in a perceptive book all the better for its brevity, while the relations between medicine and mathematics were long mooted, it was in the early nineteenth century that the balance finally tipped in favour of their marriage. A changing intellectual climate played its part. Auguste Comte's fashionable positivism offered a vision of a social science based on physical laws like any natural science, while Adolphe Quetelet was pioneering his challenging new statistical concept: average man. Objections that man could not be studied like molecules were pooh-poohed as antediluvian.

Within medicine proper, it was the French hospital environment that kindled the new fascination with quantification. Bursting with the expendable poor, the immense Paris hospitals became, in effect, great laboratories of the new scientific gaze created by pathological anatomy.

Reflecting the mathematician Poisson's notion of the 'law of large numbers' — an underwriting of probability theory — the physician Pierre Louis began systematically to tabulate medical events. Was bloodletting a beneficial therapy in diseases such as typhoid or pneumonitis? Don't rely on personal judgement, declared Louis, collect the data. He did, and found, contrary to orthodoxy, that patients subjected to venesection had a notably worse prognosis.

Louis's 'numerical method' inevitably sparked fierce controversy. Many clinicians, particularly the Montpellier professor Risueño d'Amador, turned on him with all the predictable (and not invalid) objections: no two cases are alike; reductionism ignores individual differences; clinical judgement is threatened by number-crunching. Matthews skilfully charts the disputes between Louis and his adversaries, and shows how ideas developed by Louis within pathology were later to lend

support to Claude Bernard's physiology. Bernard's aim of determining normal metabolic function implied significant assumptions about statistical probability and standard variation. He was in no doubt: "medicine is a science and not an art".

At a time when the tabulation of vital statistics was becoming routine in public health and state medicine, the Louis-Risueño d'Amador match was then, in effect, replayed on various national pitches, notably by Gustav Radicke and Karl Vierordt in Germany in the 1850s. In his most fascinating chapter, Matthews explores the joust in early twentieth-century Britain between the leading bacteriologist Almroth Wright and the pioneering biometricians Karl Pearson and Major Greenwood. In his own way Wright was a staunch progressive, full of faith in laboratory science, but when his immunological claims were challenged by statisticians he allowed himself to be manoeuvred into the fogeyish corner of declaring that he didn't understand them, but knew they must be misleading, and that he wanted nothing to do with them (similar things, one suspects, were being said about Cézanne and Schoenberg).

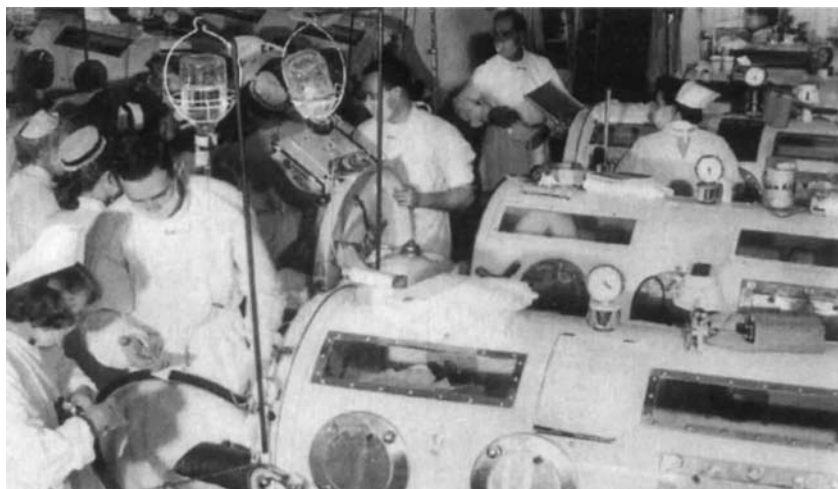
The Greenwood-Wright debate was an honourable draw, and resistance to medical quantification remained respectable at a time when fears were increasing that the triumph of scientific and technological

medicine was neglecting the sufferer and jeopardizing the doctor-patient relationship. Why, then, asks Matthews, after such a long history of reservations about the applicability of enumeration, did the clinical trials on streptomycin developed by Austin Bradford Hill in 1946 prove such an overwhelming success, finally converting the profession to a position about which it had so long remained divided? In part it was because the new clinical trials were conceptually watertight: the introduction of the double-blind element with proper controls really did, for the first time, make the numbers what they had always been claimed to be: reliably objective.

But another factor was at work, Matthews insists. With the marketing of ever more potent industrially produced drugs, and with shocking scandals such as thalidomide in the early 1960s, the public and the politicians were no longer disposed to accept 'clinical judgement' on the say-so of illustrious physicians. At long last the state required proof of the safety of medical procedures. Numbers offered a language that the politicians could trust.

This is a most illuminating study, which should be read alongside recent books by Theodore Porter, Ian Hacking and other historians who have analysed the assimilation of other aspects of the human condition within the growing 'empire of the probable'. If it is sobering to be reminded how readily in matters of life and death we relied, until so recently, on the personal judgement of fallible but often highly opinionated physicians, it should be of some comfort that today we are protected by rigorous clinical trials. □

Roy Porter is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BN, UK.



Iron lungs were — and for some they remain — life support systems for victims of poliomyelitis. This picture appears in *A Summer Plague: Polio and Its Survivors* by Tony Gould. The author, himself a survivor, provides a comprehensive account of the rise and fall of the epidemic. A compelling mix of clinical, scientific and political — as well as personal — perspectives. Yale University Press, £19.95.

Big brothers

Stuart Sutherland

The Romance of American Psychology: Political Culture in the Age of Experts. By Ellen Herman. University of California Press: 1995. Pp. 406. \$35, £28.

In the United States, psychology has become the flavour of the decade: introductory psychology is much the most popular university course and more PhDs are taken in the subject than in any other. *The Romance Of American Psychology* traces its massive growth and that of the other 'social sciences' between 1940 and 1970: readers will regret that the book does not cover the past 25 years, during which humanistic psychotherapists have imbued yet greater numbers of clients with the false idea of 'personal growth'.

The book claims that the main boost for the 'social sciences' came from the Second World War, when psychologists and psychiatrists screened potential servicemen using batteries of personality tests, most of which, particularly projective tests such as the Rorschach, were useless; they investigated enemy morale, concluding helpfully that both Germany and Japan were sick societies; they made suggestions for improving the morale of US troops, discovering to their surprise but to nobody else's that it depended on good leadership, consideration, comradeship and as much comfort as was compatible with fighting (how the British envied the GIs' rations). The investigators were surprised to find that most servicemen did not realize they were fighting for democracy. They did not even understand why the United States had entered the war, but, as Ellen Herman fails to point out, it is hard to see the relevance of this knowledge, because the declarations of war made by Japan and Germany gave the United States no choice.

During the Cold War, psychologists were employed — in a masked attempt to prevent insurgency — to investigate countries whose regimes the US government wished to support. One example is the Camelot project, which came to grief when the Chileans discovered that it was supported not, as they had been told, by the National Science Foundation, but by the US Army. Nearer home, social scientists were summoned to Washington to give advice on every kind of social problem. In 1967, immediately after the Watts riots, the Kerner commission was set up to investigate black unrest: it concluded that it was caused by the inequality, lack of opportunity, poverty and unemployment of American negroes (one of the few sensible findings made by a US social science committee). President Lyndon

New in paperback

Black Holes and the Universe by Igor Novikov. Cambridge University Press, £5.95. Novikov has been hailed as 'Russia's answer to Stephen Hawking'. The book is similar in content to *A Brief History of Time*, but written at a simpler level and illustrated with a series of cartoon drawings.

Planetary Overload: Global Environmental Change and the Health of the Human Species by A. J. McMichael. Cambridge University Press, £7.95, \$11.95. The author brings a broad evolutionary, biological, social and economic perspective to bear on the ecological disruptions threatening the human species.

The Outer Reaches of Life by John Postgate. Cambridge University Press, £6.95. A lucid popular account of the wonders of the microbial world. Reviewed by Robert Desowitz in *Nature* **368**, 820 (1994).

The Child's Path to Spoken Language by John L. Locke. Harvard University Press, £12.50. "A fascinating, scholarly and clearly written account... examining the perceptual, social, neural and cognitive precursors [of language] from before birth to the

appearance of the first recognizable words", Paul Fletcher (*Nature* **365**, 25; 1993).

A New Species of Trouble: The Human Experience of Modern Disaster by Kai Erikson. Norton, £8.95. Reviewed by W. Yule in *Nature* **372**, 140 (1994).

Autumn Books

Next week's issue will contain *Nature's* Autumn Books supplement, featuring reviews by Ziauddin Sardar (the future), Peter J. Swales (why Freud was wrong), Walter Gratzer (science as literature), Richard Davenport-Hines (history of the contraceptive pill), Peter J. Bowler (Gregor Mendel), Stuart L. Pimm (anti-environmentalism), Bryan Sykes (tracing the genes of Anastasia and Queen Victoria) and Anthony Snodgrass (Heinrich Schliemann).

Children's Books

On 16 November, *Nature's* first Children's Books supplement will be published. It will cover a range of science publications for children, ranging from a pop-up dinosaur book, through encyclopaedias of science, to a CD-ROM interactive adventure in astronomy.

Johnson promptly sacked the commissioners.

As Leonard Doob, a psychologist, remarked, his work was used only when it suited the interests of policy-makers. Nor is there any evidence that the 'social scientists' employed by government achieved anything, apart from the exclusion from the armed forces of a few delinquents and psychotics — even there they seem to have gone astray, for, according to Herman, one psychiatrist thought manic-depressives particularly suitable for military service. The advice of the Vietcong Motivation and Morale project was particularly disastrous: it concluded that the Vietcong could be conquered by global bombing, which was precisely what the government wanted to hear. As a result, 713 million leaflets were dropped on Vietnam in a month: doubtless its citizens put them to good use.

To their credit, some psychologists admitted their inability to help. Samuel Stouffer, the leading investigator of morale in the Second World War, remarked that "if the Army should ask us what single practice General Osborn's million-dollar research operation has *proved* to be helpful to morale, we honestly could not cite a scrap of scientific evidence", while Gordon Allport, the doyen of personality research, admitted that nothing was known about civilian morale. Others were more optimistic — and more self-serving. Edwin Boring announced that psychology would "enlarge the democratic communal basis of thinking", while Kubie, a respected psychiatrist, believed it could "bring the

freedom to change an entire culture". The idea that the psychotherapist could become the Big Brother of American society was widely promulgated by psychologists whose views of how to control people were wildly different; B. F. Skinner and Abraham Maslow agreed on only one point — there should be a fascism of the élite.

One theme that runs through this book is the gradual switch among psychologists from extreme right-wing attitudes towards the left. In the 1920s, they, together with psychiatrists, were proposing eugenics as a solution to America's problems and used intelligence tests based on Western culture and the American language to prevent immigration by other races. By 1970 the emphasis was on nurture not nature. Everyone was potentially good — people merely required the right upbringing — or, in Carl Rogers's opinion, the right psychotherapy — after which "change and constructive personality development will *invariably* occur" (his italics). Whether social scientists were leaders or followers in this change of political outlook is open to doubt.

The material Herman presents is fascinating, but her book has faults. It is overburdened by scholarship. Few readers will be thrilled to learn that "the Princeton Listening Center, relocated in Washington in 1941, was incorporated into the Federal Communications Commission as the Foreign Broadcast Monitoring (later Intelligence) Service, where it was directed by Goodwin Watson, a social psychologist". Moreover, despite her occasional flashes of wit (she describes the non-directive therapist Rogers as "a cheer-

leader watching the client engage in 'self-help'"), much of her prose is turgid to the point of unintelligibility — "To this book, I bring a commitment to grasping the immediacy of lived experience with genuine interest and respect, placing subjectivity within the scope of serious inquiry, and challenging experts' fervent campaign to legitimize psychology by imitating the philosophical and methodological patterns of natural science". Well, fancy that. The wording suggests she has herself been through an extensive course of psychotherapy. Although knowledgeable about the government organizations for which psychologists worked, she rarely tells us what — if anything — they found out nor to what use — if any — the government put their findings. Indeed, the main effect of the growth of clinical and social psychologists appears to have been not on society but on individual middle-class Americans whom it rendered unpleasantly selfish and unhealthily introspective. Perhaps the main mistake made by the US government was to employ psychologists rather than members of the advertising profession. After all, it is surely easier to make people change their politics than to change the brand of beans they buy. □

Stuart Sutherland is in the Laboratory for Experimental Psychology, University of Sussex, Falmer, Brighton BN1 9QG, UK.

Stasis fossilized

Laurence D. Hurst

Reinventing Darwin: The Great Evolutionary Debate. By Niles Eldredge. Weidenfeld and Nicolson: 1995. Pp. 244. £8.99.

SOME things never change. According to Niles Eldredge, species change only during speciation, but thereafter there is stasis: punctuation then equilibrium. His latest volley in this old debate fails to prove the case — no change there.

The author and his palaeontological colleagues have for the past quarter of a century challenged the conventional wisdom of evolutionary geneticists such as George Williams, Richard Dawkins and John Maynard Smith. They argue that the reported pattern of punctuation and stasis revealed in the fossil record shows that these 'defenders of the faith' have missed something. The more vigorous advocates insist that hopeful monsters may be real, that selection is overrated and that a new evolutionary synthesis is necessary (but just what this might consist is ambiguous).

Eldredge sensibly divorces himself from most of these extremes. As an apologist for Stephen Jay Gould, he also denies that many of these arguments were ever made. What is clear is that the author

takes a comparatively reasonable line: selection is of central importance, a brand-new science is not needed and the idea of hopeful monsters should be buried. But this is neither a plea for abolition from the evolutionary geneticists nor an impartial overview.

Eldredge rightly criticizes evolutionary geneticists for uncritically accepting 'just-so stories'. But whereas he denigrates these adaptationist hypotheses as mere yarns, he presents unquestioningly his own preferred theories, such as Hugh Paterson's model of speciation and Sewall Wright's analysis of structured populations: tests are not discussed or proposed.

He also fails to mention the experimental approach when discussing how to examine adaptationist hypotheses. A discussion of the correspondence of theory to data in the field of sex-ratio evolution, for example, might have allowed the uninitiated reader to appreciate the value of selectionist thinking. Evolutionary genetics instead comes across as an exclusively theoretical field redeemed only by the intervention of the author and his colleagues.

Such subtle biases make the book inappropriate for the general audience for whom it is intended, which is a pity, as the author's style is easy-going and accessible. But it cannot be so easily dismissed academically. The author has two serious points of contention. First, he asserts that stasis in the fossil record is the rule rather than the exception. But instead of details, we are given only anecdotes: some colleagues tell him that most of the time they find stasis rather than change. Second, under the assumption that stasis is real, he argues that if there are trends in the evolution of lineages, then this must be because of some higher-level process rather than simple gradual change in species morphology. He claims that evolutionary geneticists have never acknowledged the hierarchical organization of life, a perspective that, he says, his school has a monopoly on and which is necessary for a full evolutionary biology. But this is only partly true: hierarchical thinking has been an important component of evolutionary genetics since at least the mid-1960s.

What is true is that evolutionary geneticists have largely ignored what the author and his colleagues have been saying about hierarchical selection. Here then is the root of his grievance: palaeontologists have not been granted residence at the 'high table'. Part of the reason, as Eldredge explains, is that the problems in hierarchical selection examined by his school have been seen as irrelevant to debates in evolutionary genetics. But that, I suspect, is not the whole story.

Eldredge admits that the term at the heart of the issue, 'species selection', has been thrown about like a rag doll; the same is true of most of the other key



THE orchid *Brassia maculata*, from *Native Orchids of Belize* by I. McLeish, N. R. Pearce and J. S. Adams. The authors record more than 70 species previously unknown in Belize (formerly British Honduras), as well as two previously unknown to science. Balkema, £66.

words in the debate, such as emergence, constraints and contingency. With such a tangle of terminology, is it any wonder that outsiders are reluctant to enter the fray? What's more, that the uncertain terminology was (or appeared to be) used to obfuscate the debate has made it easy for many to dismiss the ideas altogether.

Eldredge claims, for instance, that his perspective of selection is different in an important respect from that of Williams and Dawkins. Organisms have, he argues, both "economic" and "reproductive" activities and this, he suggests, is missed by evolutionary genetics. But in its mathematical form, this view of selection seems to be no different from that of conventional population genetics.

Perhaps this is the great stumbling-block: the debate has been dominated by metaphor, not by mathematics. Contrast this with the equally acrimonious debate between selectionists and neutralists. Here the issues at stake are more transparent because the mathematics makes the assumptions clearer, the conclusions more robust and the disagreements open to experimental resolution.

Yet Eldredge makes, I think, an acceptable case that there really were (and are)

issues at stake as well — not great issues as the author insists, but issues nonetheless. Constructive things have come out of the slanging matches. As he discusses, the criticism that phylogeny was not taken into account in the testing of adaptive explanations provided the stimulus behind the comparative method.

Perhaps more importantly, the interventions of palaeontologists hint at difficulties to be faced by mainstream evolutionary biology. Some things have changed. Most notably, the tempo and mode of evolution are back on the agenda. Although stasis in the fossil record is still an open issue, there is a growing acceptance that other kinds of stasis constitute a real problem. Why, Williams has asked, do all eutherian mammals, whether Arctic or Saharan, have more or less the same body temperature? Differences would have been expected and, were they to be found, would be just another example of adaptation. Instead of a heated debate, is it not about time that we had a debate about heat? □

Laurence D. Hurst is in the Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK.

Karube argues that "in the developed countries where materialism has reached saturation point, the future of electronics depends on the problem of what sort of approach to take towards the brain, the neurons and the mind". On this subject, two different answers are given, even in this short book. Kazuhiro Fuchi suggests that most human intellectual activities will gradually turn into sports, whereas Shun-Ichi Amari speculates that the human brain processes an object not in terms of symbols but intuitively and unconsciously as form. He thus proposes that increasingly machines will perform symbol processing.

With such a spread of opinion, it is hardly surprising that MITI has difficulties in identifying the industries of the future. What is surprising is that of the two leading Japanese technologies, optoelectronics and production technologies, the book all but ignores the second. This is part of Japan's problem. Some 75 per cent of its exports is production equipment for components and materials. The development of these products takes a long time and, with increasing financial difficulties in Japan, their future is unclear. Yasuharu Suematsu illustrates the malaise with the comment that "the head of a securities firm said that it didn't matter if Japanese technology worked out provided Japan keeps control of finance and information. I was shocked."

This shows that Japan's problems are becoming progressively similar to those of the West and that its academics may have to follow the example of their colleagues in the United States where "you have to shout and publish often, otherwise you don't survive", as Makoto Nagao puts it. For those with a general interest and background knowledge in technology, many clues about how their colleagues in Japan think, especially those at Tokyo University, can be gained from this book. □

Anders Hansson is at 20 Leyborne Avenue, London W13 9RB, UK.

Speaking for themselves

Anders Hansson

Technology's New Horizons: Conversations with Japanese Scientists. Edited by Hiroaki Yanagida. *Oxford University Press: 1995. £19.99, \$35.*

ACADEMIC historians dislike simple models. But for those who lack historical memory, simple models provide a starting point for an understanding of the past, present and future. They are particularly useful for Japan, and can go some way to explain the pessimistic mood prevalent there today. Beginning with the Meiji restoration, Japan's history can be divided into 40-year cycles: 1865–1905, from the restoration to victory over Russia; 1905–45, from increasing militaristic expansion to defeat; and 1945–85, with the rebirth of economic power.

If the trend continues, then the present cycle, lasting to 2025, will be one of decline. True, export-led reinvestment following the 1973 oil-price crisis did result in an annual growth of 6 per cent in the late 1980s. But many economists, industrialists and politicians during that period, including Nakasone Yasuhiro, the prime minister from 1982 to 1987, were in fact principally preoccupied with trying to avert the beginning of the fourth cycle of economic decline. As it turned out, 1986 was the high point of Yasuhiro's time in office, and in 1990 the bubble

burst. The effects continue to undermine the economy today. Again, the reason is largely historical: for many senior executives it is vital to avoid what happened in the Showa depression (1927–32), which resulted in the active militarism that still tarnishes Japan's international reputation.

So an industrial revival without an economic crisis like that of Anglo-American capitalism is the aim of most Japanese leaders. The Ministry of International Trade and Industry (MITI) assumes that this can be achieved along with an annual economic growth of 1 per cent, which gives time for change. Industrial horizons are closer than that. Within the next five years, new industries must be developed to maintain employment and social stability, leading industrialists claim.

Against this general background, *Technology's New Horizons* gives an interesting insight into the views of senior scientists in Japan. As usual, it is often the contributors' off-beat comments that reveal the most important things. The editor, Hiroyuki Yoshikawa, wonders "if capitalism has the power to resolve the contradiction between these infinite desires and finite resources". He even talks about an "infinity cult" in the natural sciences: "biology may have more of a flavour of what is finite, but molecular biology, I think, has no interest in it at all". Isao

Corrections

In this year's New Journals supplement (21 September), the caption on p.265 refers to a book by Christopher Langton as *Artificial Intelligence*. It is in fact entitled *Artificial Life: An Overview*. MIT Press, \$42, £24.95.

The wrong publisher was given for three books recently listed as new in paperback. *Science on Trial* by Douglas J. Futuyma, *A Primer in Ecology* by Nicholas J. Gotelli and *Exploring Evolutionary Biology* edited by M. Slatkin are published by Sinauer and distributed in Europe by W. H. Freeman.

Vascular system defects and neuronal apoptosis in mice lacking Ras GTPase-activating protein

Mark Henkemeyer^{*§}, Derrick J. Rossi^{*†}, Douglas P. Holmyard^{*},
Mira C. Puri^{*†}, Geraldine Mbamalu^{*}, Kendraprasad Harpal^{*},
T. Shane Shih[‡], Tyler Jacks[‡] & Tony Pawson^{*†}

^{*}Programme in Molecular Biology and Cancer, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada

[†]Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada

[‡]Howard Hughes Medical Institute, Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The gene encoding p120-rasGAP, a negative regulator of Ras, has been disrupted in mice. This *Gap* mutation affects the ability of endothelial cells to organize into a highly vascularized network and results in extensive neuronal cell death. Mutations in the *Gap* and *Nf1* genes have a synergistic effect, such that embryos homozygous for mutations in both genes show an exacerbated *Gap* phenotype. Thus rasGAP and neurofibromin act together to regulate Ras activity during embryonic development.

THE Ras family of small guanine-nucleotide-binding proteins are conserved molecules thought to be involved in a diverse array of normal and oncogenic signalling pathways¹⁻³. Several regulatory proteins modulate Ras activity in normal cells⁴. Guanine-nucleotide exchange factors activate Ras by promoting the exchange of Ras-bound GDP for GTP, thereby inducing the association of Ras with several potential downstream signalling molecules, such as the Raf protein kinase and phosphatidylinositol-3' kinase^{3,5}. In contrast, GTPase-activating proteins (GAPs) stimulate the intrinsic rate of Ras GTPase activity, and therefore function as negative regulators by accelerating the conversion of active GTP-bound Ras to the inactive GDP-bound state⁶. Four GAPs specific for Ras proteins have been identified in mammalian cells, these being p120-rasGAP (rasGAP), neurofibromin (the product of the *Nf1* tumour-suppressor gene), and two members of the GAP1 family⁷⁻¹².

rasGAP is a widely expressed modular protein composed of a carboxy-terminal GAP domain that specifically recognizes GTP-bound Ras, a central pleckstrin homology (PH) domain, and two amino-terminal Src-homology 2 (SH2) domains that flank a single SH3 domain¹³. The presence of SH2 domains suggests that rasGAP regulates Ras proteins downstream of protein tyrosine kinases (PTK). Indeed, rasGAP is phosphorylated by a variety of activated PTKs, and associates through its SH2 domains with autophosphorylated receptors and two intracellular proteins, p62 and p190 (refs 14-19), which contain phosphotyrosine (pTyr). One of these proteins, p190, contains a GAP domain specific for Rho-related GTPases, which direct the formation of actin stress fibres and focal adhesions^{20,21}. The ability of rasGAP to bind multiple proteins through its various domains suggests that it might be involved in several biological processes²².

To define the biological functions of rasGAP and investigate the potential consequence of Ras activation in mammalian development, we have disrupted the gene encoding rasGAP (*Gap*) and analysed the effects of this mutation on mouse development. By combining the *Gap* mutation with a previously described²³ mutation in the gene encoding murine neurofibromin (*Nf1*), we have also assessed the relative roles of rasGAP and neurofibro-

min in controlling the growth and development of early post-implantation embryos.

Embryonic stem cell growth

The coding region for the N-terminal rasGAP SH2 domain was deleted by homologous recombination in embryonic stem (ES) cells (Fig. 1a, b). *Gap* +/− ES cells were subsequently plated in a high concentration of G418 to induce selection of *Gap* −/− cells through duplication of the mutant allele, which contains a neomycin resistance cassette. Most of these cell lines (33 of 51) had lost the wild-type *Gap* allele, and western blot analysis using two different rasGAP-specific antibodies showed that they were deficient for expression of rasGAP (Fig. 1c). The *in vitro* growth properties of *Gap* −/− cell lines were similar to those observed for *Gap* +/− cells (data not shown). These results indicate that the *Gap* mutation generates a null allele, and that rasGAP is not crucial for the viability or proliferation of ES cells.

Embryonic development

Chimaeric mice were generated with a +/− ES cell line, and germline transmission of the targeted allele was obtained. The *Gap* mutation was crossed into an inbred (129/Sv) background and into two other backgrounds (C57Bl/6 and CD1). *Gap* +/− mice appeared normal and showed no increase in mortality or tumour incidence over an 18-month period. However, no *Gap* −/− homozygous animals were born following intercrosses between heterozygous males and females, indicating a recessive lethal phenotype. *Gap* −/− homozygous embryos (henceforth referred to as *Gap* mutants) appeared visibly abnormal by embryonic day 9.25 (E9.25) and subsequently died by E10.5. The phenotypes associated with the *Gap* mutation were indistinguishable on the different genetic backgrounds.

Gap mutants dissected at E8.5 (6-12 somites), at a stage in murine development called turning, appeared morphologically normal, and were identified by either Southern blot analysis (Fig. 1b) or *Gap*-specific polymerase chain reaction (PCR) (Fig. 1d). At this stage of development rasGAP was easily detected in immunoblots of *Gap* ++ and +/−, but not −/−, embryos (Fig. 1e). By E9.25 (18-20 somites), *Gap* mutants became visibly identifiable as they had a defect in posterior elongation, which by E9.5 resulted in developmental arrest with less than

§ To whom correspondence should be addressed.

25 somites (Fig. 2a). No abnormal or excessive cell proliferation was observed in *Gap* mutants. Indeed *Gap* mutants at E9.5 were significantly smaller in size than their littermates, and had obvious defects in the development of the vascular system.

Vascular system defects in *Gap* mutants

The embryonic circulatory system comprises both extra-embryonic (yolk sac) and embryonic vascular networks, which together ensure that a supply of circulating blood is provided to the developing embryo. The yolk sac is the source of early embryonic blood, and is composed of an outer endoderm layer and an inner mesoderm layer. Development of the yolk-sac vasculature begins with the formation of mesodermal blood islands containing haematopoietic stem cells surrounded by endothelial cells²⁴. In early vasculogenesis these blood islands fuse, forming an organized, honeycombed array of endothelial cells. Between E8.5 and E9.5 the endothelial cells again alter their migration and adherence pattern as they reorganize into a highly vascularized network that feeds into the embryo itself. At this stage,

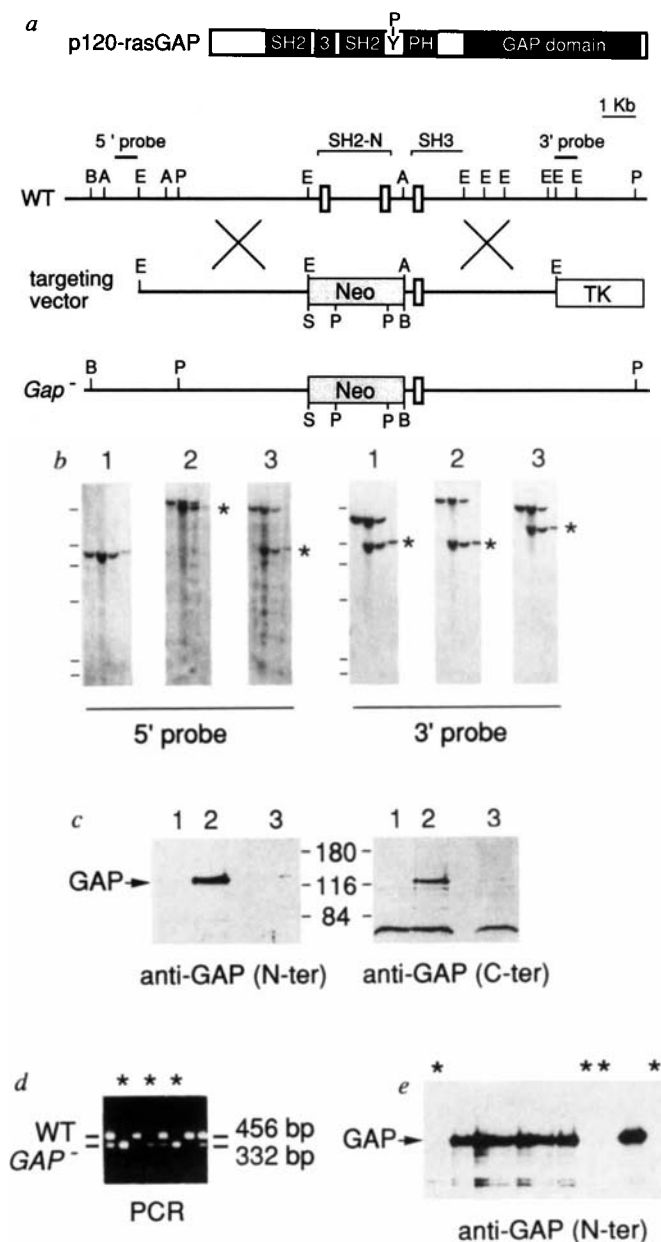
endothelial cells within the embryo have formed a vascular network including the heart and major blood vessels.

The yolk sacs of *Gap* mutants were visibly abnormal by E9.5, exhibiting a roughened or wrinkled appearance when compared with the smooth yolk sac of normal embryos. To visualize this yolk-sac phenotype, a transgene²⁵ that expresses *lacZ* specifically in endothelial cells (*Tek-lacZ*) was crossed into the *Gap* mutant background, and embryos were collected and stained for β -galactosidase (β -gal) activity. In whole mounts of stained E10.5 yolk sacs it was clear that the reorganization of endothelial cells in *Gap* mutants was delayed when compared to wild-type littermates (Fig. 2b). In the mutants, endothelial cells aggregated into a honeycombed pattern but subsequently failed to reorganize into a vascular network. Stained sections of E9.5 *Tek-lacZ* yolk sacs showed that the endothelial cells of *Gap* $+/-$ heterozygotes formed major blood vessels, whereas *Gap* mutants formed abnormal chambers containing blood cells (Fig. 2c, d).

The *Tek-lacZ* marker was also used to visualize early endothelial cell abnormalities within mutant embryos. To investigate

FIG. 1 Targeted mutation in *Gap*. a, Protein structure, genomic restriction map and targeting strategy used to delete a 2.4-kb region of *Gap* containing the coding exons for the rasGAP N-terminal SH2 domain. Locations of the SH2, SH3 (3), PH and GAP domains, as well as the tyrosine phosphorylation site (P-Y), are indicated on the structure of rasGAP. Three exons on the wild-type (WT) murine *Gap* gene have been sequenced and are indicated by boxes. The external 5' and 3' probes used to identify homologous recombination events are indicated above the WT restriction map. Only pertinent restriction sites are indicated in the maps of the targeting vector and the *Gap* mutation (*Gap*⁻). A, *Asp*718; B, *Bam*HI; E, *Eco*RI; P, *Pst*I; S, *Sal*I. b, Southern blot analysis of genomic DNA digested with *Pst*I (1), *Sac*I plus *Sal*I (2), and *Bam*HI (3) were hybridized with either the 5' or 3' *Gap* probes. Each panel contains, from left to right, lanes of wild-type 129 tail DNA, *Gap* $+/-$ tail DNA, *Gap* $+/-$ ES cell DNA, and *Gap* $-/-$ yolk-sac DNA obtained from a mutant at E9.5. Asterisks mark the positions of mutant bands and the migration of Lambda DNA digested with *Hind*III is indicated. c, Western blot analysis of total protein lysates from two *Gap* $-/-$ ES cell lines (lanes 1 and 3) and one *Gap* $+/-$ cell line (lane 2) was performed with two different anti-rasGAP antibodies. The position of p120-rasGAP is indicated. d, PCR analysis of yolk-sac DNA obtained from the progeny of a heterozygous intercross dissected at E9.0. The position of wild-type (WT) and mutant (*Gap*⁻) PCR products is indicated. Asterisks denote the lanes where *Gap* $-/-$ samples were loaded. e, Western blot analysis of total embryo protein lysates from the progeny of a heterozygous intercross dissected at E8.5. The position of p120 rasGAP is indicated by an arrow, and asterisks denote the lanes where *Gap* $-/-$ samples were loaded.

METHODS. The *Gap* targeting vector was constructed by inserting 4.8 kb *Eco*RI and 4.7 kb *Asp*718-*Eco*RI fragments of cloned 129 strain *Gap* genomic DNA into pPNT⁴⁶. A homologous recombination event inserts a neomycin resistance cassette while deleting 2.4 kb of the *Gap* locus including two exons encoding the rasGAP N-terminal SH2 domain (codons 172–221 and 222–267 of the published⁴⁷ murine rasGAP cDNA sequence). The DNA sequence of the deleted exons (and the adjacent exon encoding amino acids 268–290) indicates that any aberrant splicing around the neomycin resistance cassette would generate a defective *Gap* transcript containing a frameshift mutation. The linearized targeting vector was electroporated⁴⁸ into the ES cell line R1 (ref. 49), and colonies were isolated following G418 and gancyclovir selection. Addition of gancyclovir resulted in a 6.7-fold enrichment. *Pst*I digests of genomic DNA from 108 double-resistant ES cell clones were hybridized with an external 3' *Gap* probe identifying 8 integration events. Additional restriction digests and a 5' external probe verified that all 8 clones contained a homologous recombination into *Gap*. Germline-transmitting chimaeras of one cell line were obtained by standard injection into C57B1/6 blastocysts, and the mutation was crossed into various genetic backgrounds as described in the text. The *Gap* locus was routinely genotyped using the following three oligonucleotides: PCR*Gap*1, 5'-CGGCCTACGATCACTCTCTTATAAG-3'; PCR*Gap*2, 5'-CTTGCTCTTATGTTCCAGGAAAGG-3'; and PCR*Neo*2, 5'-TACCCGGTAGAATTGACCTGAG-3'. Total cell (c) or total embryo (e) protein lysates were subjected to western blot analysis using polyclonal antibodies¹⁶ directed against the N-terminal SH2 domain of rasGAP. Western blots were developed using chemiluminescence. The filter in c was reprobed with polyclonal antibodies that recognize C-terminal epitopes of rasGAP



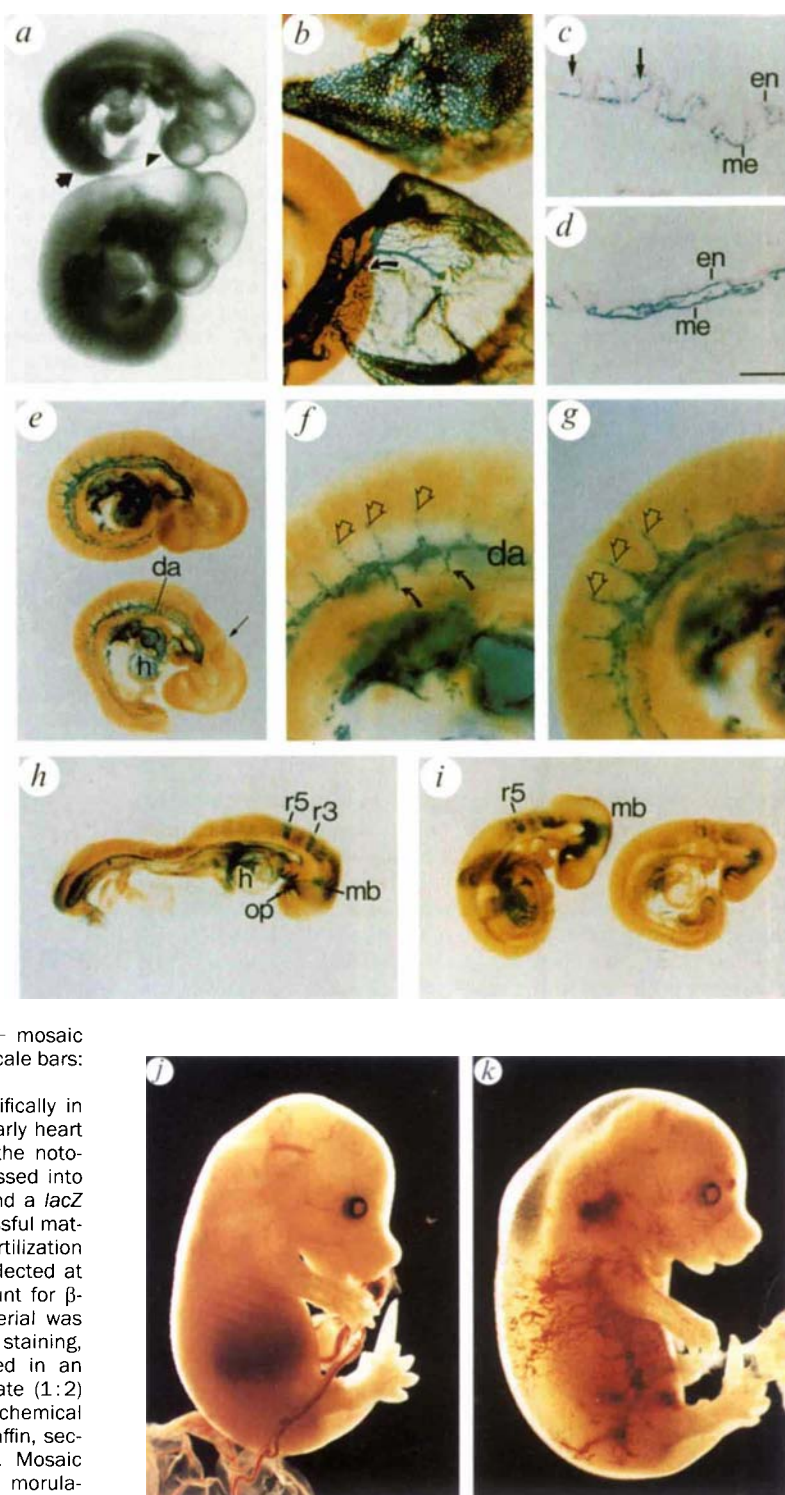
(ONYX Pharmaceuticals). Equal amounts of cell lysate were loaded on the gel as indicated by the background band visible below *M*, 84 K in all three lanes on the anti-GAP (C-ter) blot.

early vascular phenotypes we analysed mutants with 8–18 somites. At this early stage the mutants do not exhibit any overt phenotype and are still of comparable size to heterozygous and wild-type littermates. As judged by β -gal staining, *Gap* mutants with up to 12 somites formed an apparently normal endothelial cell architecture, including a chambered heart and major blood vessels (not shown). At the 16-somite stage, the dorsal aorta of *Gap* mutants exhibited correct branching to form

intersegmental arteries that project dorsally between the somites (Fig. 2e). However, closer examination showed that *Gap* mutants exhibited a thinning of the dorsal aorta and aberrant ventral branches that resemble intersegmental arteries (Fig. 2f). These irregular, ventral-projecting endothelial cells were not observed in heterozygous embryos (Fig. 2g). In late-stage *Gap* mutants (E9.5) the embryonic vasculature exhibited local ruptures causing leakage of blood into the body cavity, and the heart became

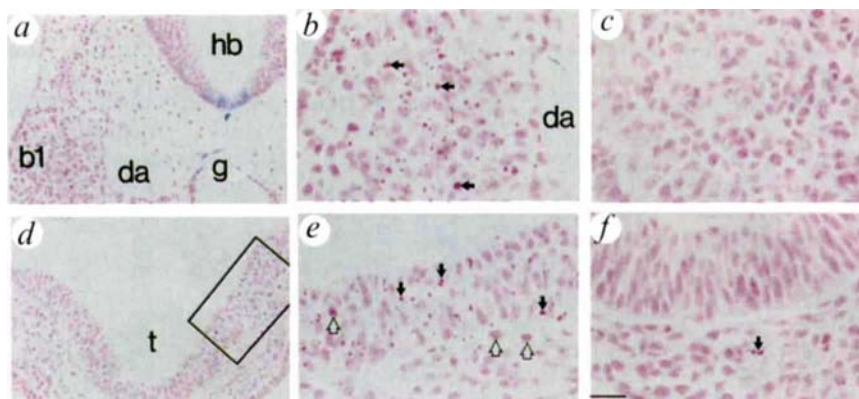
FIG. 2 *Gap* mutant phenotype. For the embryos, dorsal is up and anterior is right, unless otherwise indicated. **a**, Whole mount of a *Gap* $-/-$ homozygote (top) and normal littermate (bottom) isolated at E9.5 showing the reduced size of the mutant. The pericardial sac of the mutant was distended (arrowhead) and posterior elongation had arrested with fewer than 25 somites (arrow). **b–g**, Endothelial cells in *Gap* mutants were stained with a *Tek-lacZ* marker²⁵ and viewed as whole mounts (**b**, **e–g**) or tissue sections (**c**, **d**). **b**, Blue-stained endothelial cells in a mutant E10.5 yolk sac (top) exhibit a honeycombed pattern, whereas, in a normal littermate (bottom), the pattern is highly vascular with the establishment of a primary network to the embryo (arrow). **c**, **d**, Blue-stained endothelial cells in a mutant E9.5 yolk sac (**c**) form abnormal chambers, some of which contain blood cells (arrow). The yolk sac of a heterozygous littermate appears normal (**d**). Note the apparent excess of endoderm (en) over mesoderm (me) cells in the mutant yolk sac. **e**, Stained endothelial cells in E9.0 *Gap* heterozygous (top) and homozygous (bottom) whole-mount embryos. A fairly normal endothelial cell architecture was formed in the 16–18-somite mutant with a chambered heart (**h**) and major blood vessels visible, including the aortic arch and dorsal aorta (**da**). **f**, **g**, Higher magnifications of the mutant and normal 16–18-somite stage embryos shown in **e**, respectively. The dorsal aorta in both embryos exhibited correct branching of intersegmental arteries between the somites (open arrows). In the mutant the dorsal aorta was thinner and formed a number of aberrant ventral-projecting branches (curved filled arrows). **h**, The expression of a *lacZ* marker for anterior neural tube segmentation and early heart development (*Nuk^{lacZ}*) appeared normal in turning E8.75 *Gap* mutants with 10 somites. The heart (**h**), optic vesicle (**op**), ventral midbrain (**mb**), and hindbrain rhombomeres **r3** and **r5** of the developing neural tube were stained by this marker. *Nuk^{lacZ}* expression appeared normal in *Gap* mutants with up to 18 somites (not shown). However, by E9.5 (**i**), expression of *Nuk^{lacZ}* in the developing brain was significantly reduced in a *Gap* mutant (right) when compared to a heterozygous littermate (left). This reduced expression may be due to increased neuronal cell death in *Gap* mutants. **j**, **k**, E15 mosaics composed of wild-type and either *Gap* $+/-$ (**j**) or *Gap* $-/-$ (**k**) cells. The *Gap* $-/-$ mosaic exhibited severe oedema and grossly abnormal vasculature. Scale bars: **c**, **d**, **f**, **g**, 100 μ m; **e**, 400 μ m.

METHODS. Transgenic *lacZ* markers that express β -gal specifically in endothelial cells (*Tek-LacZ*)²⁵, the anterior neural tube and early heart (*Nuk^{lacZ}*)²⁶ (M.H. and T.P., manuscript in preparation), and the notochord, ventral neural tube and gut (*Cordon-bleu*)²⁷ were crossed into the *Gap* mutant background. Males heterozygous for *Gap* and a *lacZ* marker were crossed to *Gap* heterozygous females and successful matings identified by presence of a vaginal plug the next day. Fertilization (E0.0) was assumed to occur at midnight. Embryos were collected at the indicated period of development and stained whole-mount for β -gal activity as described²⁷. A small portion of yolk-sac material was used to genotype each specimen using PCR. Following β -gal staining, whole-mount tissues were postfixed in formalin, dehydrated in an ethanol series, and cleared in benzyl alcohol:benzyl benzoate (1:2) immediately before observation and photography. For histochemical analysis, tissues were returned to ethanol, embedded in paraffin, sectioned at 6 μ m, and counterstained with nuclear-fast red. Mosaic embryos were generated by aggregating one CD1 wild-type morula-staged (8-cell) embryo with a similar-staged embryo derived from a cross of inbred 129 *Gap* $+/-$ males and females. Following overnight incubation *in vitro*, chimaeric blastocysts were transferred into pseudo-pregnant females and allowed to develop *in utero* until E15. Determining the contribution of CD1 and 129 cells in these mosaics was aided



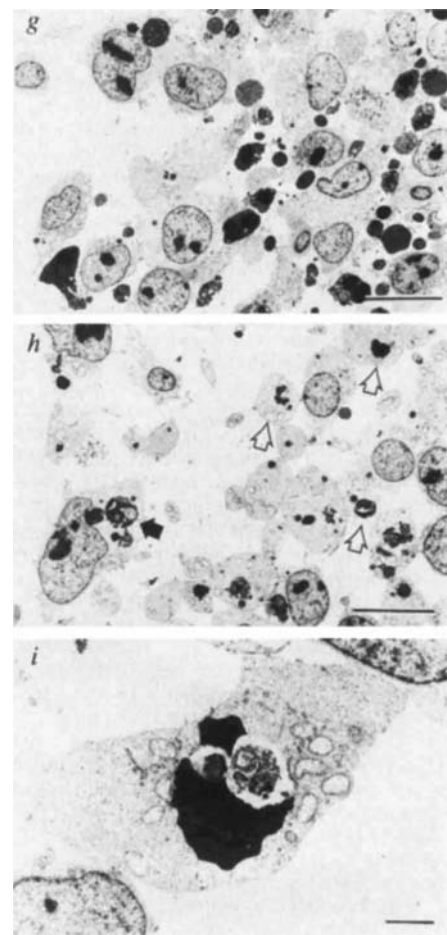
by a polymorphism of a *Pst*I site in the wild-type *Gap* allele. Brightfield photography was performed with either a Wild M10 microscope or a Leitz DMRXE compound microscope using Kodak EPY 64 Tungsten slide film or Kodak Ektar 100 print film.

FIG. 3 Cell death in *Gap* mutants. *a–f*, Transverse paraffin sections through 12–14-somite *Gap* homozygous (*a, b, d, e*) or heterozygous (*c, f*) littermates were stained with nuclear-fast red. *b, e*, High-magnification views of *a* and *d*, respectively. In *Gap* mutants, this stain intensely labelled pyknotic nuclei of dead or dying cells specifically in the hindbrain (*a*), in branchial arch 1 (*a, b*), and in more anterior regions of the brain including the telencephalon and optic stalk (*d, e*). Filled arrows denote examples of dead cells. Very few pyknotic nuclei were evident in similar sections through branchial arch 1 (*c*) and the telencephalon/optic stalk (*f*) of the heterozygous embryo. Although *Gap* mutants may exhibit a slight increase in pyknotic nuclei within branchial arch 2 (not shown), sections through branchial arch 1 and the anterior neural tube consistently revealed very high levels of cell death. The *Gap* mutants do not suffer from a general cell cycle arrest, as evident by the presence of metaphase chromosomes, indicative of cells undergoing mitosis (open arrows in *e*). Abbreviations: b1, branchial arch 1; da, dorsal aorta; g, gut; hb, hindbrain; t, telencephalon.



g–i, Transmission electron micrographs of cells within branchial arch 1 (*g*) and the telencephalon/optic stalk (*h, i*) of 12–14-somite *Gap* mutants. Numerous apoptotic cells and cell fragments were easily identified by the appearance of electron-dense chromatin masses. In *g*, shrinkage and darkening of the cytoplasm was prominent in many of the dying cells. In *h*, a large number of apoptotic bodies were observed that contained fragmented chromatin particles and intact organelles, including mitochondria (examples denoted by open arrows). A heterophagosome containing engulfed cell particles is also apparent in this micrograph (filled arrow). Viable cells were also present in these micrographs, as evident by their apparently normal nuclei. The high-magnification view in *i* shows a cell in early apoptosis, as judged by the convoluted nuclear membrane and partly degraded, electron-dense chromatin. Scale bars: *a, d*, 50 μ m; *b, c, e, f*, 20 μ m (scale bar in *f*); *g, h*, 10 μ m; *i*, 2 μ m.

METHODS. Transverse sections were obtained at the level indicated by the arrow in Fig. 2e. These sections bisect rhombomere 2 and branchial arch 1. *a–f*, Embryos at the 12–14-somite stage were collected in PBS containing magnesium and calcium, rapidly fixed, and stained for *Cordon-bleu lacZ* activity²⁷. Embryos were then embedded in paraffin, sectioned (6 μ m), and counterstained with nuclear-fast red. *Cordon-bleu*-specific *lacZ* staining of ventral structures including the hindbrain, the underlying notochord, and the gut can be observed in *a*. Photography was performed under brightfield conditions with a Leitz DMRXE compound microscope using Kodak Ektar 25 print film. *g–i*, Ultrathin (100 nm) plastic sections were stained with uranyl acetate and lead citrate before being viewed at 60 kV in a Philips CM100 Biotwin TEM.



surrounded by a distended pericardium (Fig. 2a and not shown). Observation of living late-stage mutants revealed a laborious heartbeat accompanied by a much reduced circulation of blood into the embryo. The specific effect of the *Gap* mutation on the embryonic vascular system is highlighted by the relatively normal expression patterns observed for markers of other cell types, including *Nuk*²⁶, which is expressed in specific segments of the developing brain and in the early heart (Fig. 2h, i), and *Cordon-bleu*²⁷, which is expressed in the notochord, ventral neural tube, and gut (Fig. 3a). These results indicate abnormal development of the circulatory system is likely to be a major cause of the lethality observed in *Gap* mutants.

We also generated mosaic embryos by aggregating wild-type and *Gap* $-/-$ morula-stage (8-cell) embryos. When transferred into pseudopregnant females these chimaeric embryos were allowed to develop *in utero*. Southern blotting showed that these mosaics had approximately equal numbers of wild-type and *Gap* $-/-$ cells (data not shown). The presence of wild-type cells in these mosaics allowed embryonic development to proceed beyond the typical stage of *Gap* lethality. Consistent with the early effect of the *Gap* mutation on vascular development described above, *Gap* $-/-$ mosaics examined at E15 exhibited striking defects associated with the vascular system (Fig. 2j, k). These data indicate that rasGAP acts in a cell-autonomous fashion, as anticipated from its role as a cytoplasmic signalling protein.

Cell death in *Gap* mutant embryos

Because of its GAP activity towards Ras, we analysed several mutant embryos for examples of deregulated cell growth. Conventional histochemical analysis of tissue sections from numer-

ous *Gap* mutant embryos failed to detect any regions of increased cell growth or accumulation. Unexpectedly, we observed localized regions of extensive cell death visible even in early E9.0 mutants. Densely stained pyknotic nuclei, which are indicative of dead or dying cells, were specifically observed in the first branchial arch and hindbrain (Fig. 3a–c), and in more anterior brain structures including the optic stalk and telencephalon (Fig. 3d–f). Other tissues appeared relatively free of pyknotic nuclei.

Transmission electron microscopy was used to determine whether the cell death observed in early-stage *Gap* mutants was due to apoptosis or general cell necrosis. This analysis identified a large number of apoptotic cells in the *Gap* mutants, and examples are shown for the neural crest of branchial arch 1 (Fig. 3g) and the optic stalk (Fig. 3h, i). Several characteristics of apoptosis were identified, including the appearance of electron-dense chromatin masses, convolution of the nuclear membrane,

TABLE 1 Genotype analysis of embryos generated from compound heterozygous mice

Gap	Nf1	Number	Observed (%)	Expected (%)
+/+	+/+	7	4.3	6.25
+/+	+/-	21	12.9	12.5
+/+	-/-	10	6.1	6.25
+/-	+/+	21	12.9	12.5
+/-	+/-	41	25.2	25
+/-	-/-	20	12.3	12.5
-/-	+/+	11	6.7	6.25
-/-	+/-	23	14.1	12.5
-/-	-/-	9*	5.5	6.25
Total		163	100.0	100.0

Embryos genotyped as *Gap* $-/-$; *Nf1* $+/-$ exhibited phenotypes indistinguishable from *Gap* $-/-$; *Nf1* $+/+$ littermates.

*These embryos exhibited a more severe phenotype, as described in the text.

cell shrinkage, and darkening of the cytoplasm. We also observed cell remnants, known as apoptotic bodies, containing intact organelles and fragmented chromatin particles. Many of the dying cells and cell fragments were already in a late phase of apoptosis, having been phagocytosed by neighbouring cells.

Genetic interaction between *Gap* and *Nf1*

Neurofibromin and rasGAP contain related GTPase-activating domains, which function *in vitro* to negatively regulate Ras proteins. In mice, *Nf1* $-/-$ homozygotes appear relatively normal at E12.5, but exhibit defects in heart development^{23,28} and die between E13.5 and E14.5. To determine whether mutations in

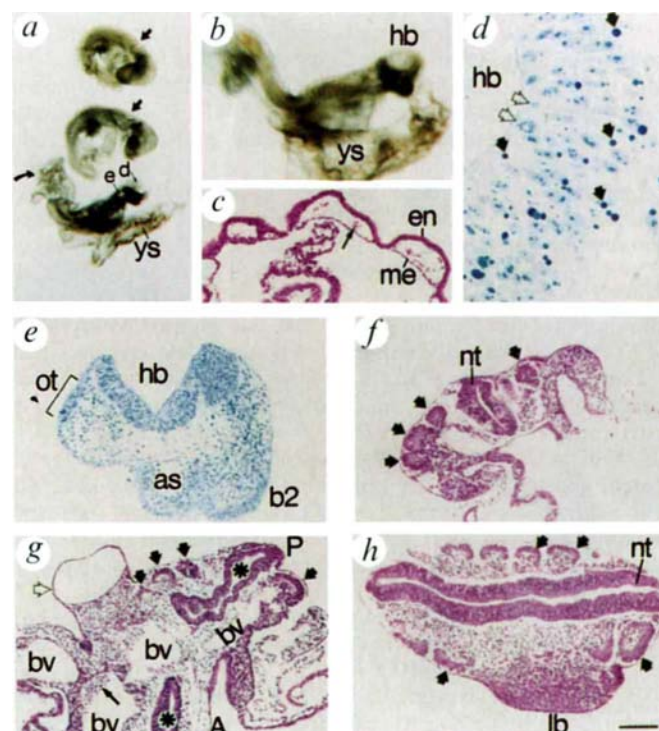
Nf1 would affect the earlier phenotypes exhibited by *Gap* mutants, adult mice heterozygous for both *Gap* and *Nf1* were generated. Following matings of *Gap* $+/-$; *Nf1* $+/-$ compound heterozygotes, embryos were dissected at E9.0 and scored for unusual phenotypes. Of 163 embryos collected, 9 had arrested development near E8.5 and exhibited more severe phenotypes than were observed for *Gap* single mutants. Moreover, these 9 embryos were identified by genotyping as *Gap* $-/-$; *Nf1* $-/-$ double homozygotes and accounted for all of the expected 1/16 of the total progeny class (Table 1). In contrast, embryos homozygous for both *Gap* and another lethal mutation, *Notch1* (ref. 29), did not exhibit enhanced phenotypes as observed for *Gap*; *Nf1* double homozygotes (M.H. and R. A. Conlon, unpublished data). These results demonstrate a level of specificity for the genetic interaction between *Gap* and *Nf1*.

Inspection of whole-mount *Gap*; *Nf1* double homozygotes indicated that 8 of the 9 mutants failed to turn, had fewer than 8 disorganized somites, lacked limb buds, displayed an enlarged allantois, and exhibited major abnormalities in the heart and surrounding trunk (Fig. 4a, b). The ruffled yolk sac was similar to *Gap* single mutants and contained primitive, nucleated blood cells (Fig. 4c). Anterior structures, including the brain, optic and otic vesicles and the branchial arches, were either absent or abnormal. The remaining double mutant exhibited a more severe phenotype and had arrested development near E7.5 (not shown).

Thorough histological analysis of four *Gap*; *Nf1* double homozygotes indicated abnormal development throughout the embryo. Sections through the anterior neural tube revealed that three of four specimens failed to develop any recognizable mid-brain and forebrain structures. A rudimentary anterior brain was visible in the remaining embryo but it had arrested development

FIG. 4 *Gap* $-/-$; *Nf1* $-/-$ double-mutant phenotype. a, b, Whole mounts of wild-type (top), *Gap* single mutant (middle), and *Gap*; *Nf1* double mutant (bottom) littermates isolated at E9.0. b, Higher magnification of a second *Gap*; *Nf1* double-mutant specimen. The posterior tails of double mutants failed to turn and the embryos arrested development with fewer than 8 disorganized somites. The double mutants exhibited a ruffled yolk sac (ys), enlarged allantois (curved arrow in a), and major abnormalities in the heart and surrounding trunk region. The dorsal surface of the hindbrain (hb) failed to close and branchial arches were abnormal or not present. Other anterior structures did not form in the double homozygotes, including the otic vesicles, midbrain or forebrain. By this stage, the wild-type and *Gap* single mutant littermates had formed 16–18 somites and developed other normal embryonic structures, including otic vesicles (arrows). c, Paraffin section of a *Gap*; *Nf1* double homozygous yolk sac showing blood cells (arrow) within the abnormal ruffled chambers. Abbreviations: en, endoderm; me, mesoderm. d, e, Transverse semi-thin plastic sections through the hindbrain of the *Gap*; *Nf1* double mutant embryo shown in a. In d, a high magnification of the anterior hindbrain revealed numerous darkly stained apoptotic cells in the neuroectoderm (filled arrows). Some cells in the brain remained viable as judged by the presence of metaphase chromosomes in the sections (open arrows). e, Low magnification of a more posterior section through presumptive hindbrain rhombomeres r4/r5. The dorsal surface of the double mutant hindbrain failed to close. Similarly, an otic placode (ot) is visible but it did not invaginate to form an otic vesicle. An abnormal branchial arch 2 (b2) and aortic sac (as) were also apparent in this section. f–h, Paraffin sections through two different *Gap*; *Nf1* double mutants (f, g) or a *Gap* single mutant (h). Compared to wild-type and *Gap* single mutants, the somites in double mutants were disorganized and often clustered to lateral regions of the embryo (thick filled arrows). In some instances (f) only condensed presomitic mesoderm was evident. The double mutant shown in g exhibited more severe defects (A, anterior; P, posterior). The posterior neural tube (asterisk) was kinked, abnormally large blood vessels (bv) containing pools of blood cells (thin filled arrows), and an unusual fluid-filled vesicle had formed (open arrow); lb, limb bud. Scale bars: c, e–h, 100 μ m; d, 20 μ m.

METHODS. Adult mice of the genotype *Gap* $+/-$; *Nf1* $+/-$ were generated and these were intercrossed to obtain *Gap* $-/-$; *Nf1* $-/-$ double homozygotes. Resulting embryos were dissected at E9.0 and rapidly fixed in 4% fresh paraformaldehyde and 1% glutaraldehyde in 0.1 M phosphate buffered saline. Embryos were washed and photo-



graphed as whole mounts before plastic (d, e) or paraffin (c, f–h) embedding and sectioning. The genotype of embryos shown in a are: *Gap* $+/-$; *Nf1* $+/-$ (top), *Gap* $-/-$; *Nf1* $+/-$ (middle) and *Gap* $-/-$; *Nf1* $-/-$ (bottom). The levels of the sections shown in d and e are indicated on the whole-mount photograph of the same embryo in a. The section in e is slightly off angle, such that the left side of the neural tube is through hindbrain rhombomere 5 (r5) and the right side is through rhombomere 4 (r4). This was determined by the position of the otic placode (ot) on the left and an abnormal branchial arch 2 (b2) on the right.

before the formation of optic vesicles (not shown). As observed in *Gap* single mutants, widespread cell death was detected in the hindbrain (Fig. 4d) and migrating cranial neural crest. In all four double homozygotes examined, the dorsal surface of the hindbrain failed to close, and otic placodes developed but none had invaginated to form vesicles (Fig. 4e). The embryonic mesoderm was also affected, resulting in a general disorganized trunk morphology and abnormal somite development (Fig. 4f-h). The cardiovascular system was more severely affected in the double mutants. In some instances the embryonic blood vessels appeared much larger, occupying more space than normal (see Fig. 4g).

A striking difference between *Gap* single mutants and *Gap*; *Nf1* double mutants was observed in the hindbrain neuroepithelium. In all four double mutants analysed, one or more localized regions of cell outgrowth, indicative of increased proliferation or survival, were evident within the neural tube (Fig. 5). These extra neuronal cells formed asymmetrical ventral-lateral bulges that invaded the underlying head mesenchyme. These abnormal growths appeared to contain up to twice as many cells as other regions of the neural tube that were not affected in the double mutants. This phenotype is specific for *Gap*; *Nf1* double mutants (see for instance an *Nf1* single mutant, Fig. 5d). These data indicate that cell proliferation pathways within the neuroepithelium might be highly sensitive to the combined absence of functional rasGAP and neurofibromin.

Discussion

The biological functions of Ras signalling in early mammalian development are unknown. Eliminating expression of GAPs with targeted recessive mutations might be expected to increase the overall level of GTP-bound Ras in the mutant embryos. Indeed, we have established *Gap* mutant cell lines from E9.5 embryos and found that they are more sensitive than wild-type cells to growth-factor-induced changes in Ras-GTP levels. (P. van der Geer, M.H. and T.P., unpublished data). In contrast, *Nf1* mutant Schwann cells respond normally to growth-factor stimulation but exhibit a higher basal level of Ras-GTP⁵⁰. These results suggest that rasGAP functions to downregulate Ras molecules that have been activated by receptor PTKs. Based on the specific defects in *Gap* single mutants, and the genetic interaction observed in *Gap*; *Nf1* double mutants, we suggest that the embryonic phenotypes observed in these animals are due, at least in part, to increased levels of Ras-GTP, and hence to aberrant Ras signalling to downstream effector pathways.

Embryos lacking rasGAP exhibited major defects in endothelial cell organization. A likely explanation is that the movement of endothelial cells, or remodelling of vascular processes, is in part controlled by receptor PTKs functioning through rasGAP to regulate the level of GTP-bound Ras. This is consistent with recent genetic studies demonstrating essential functions for the Flt-1, Flk-1, Tie-2 (Tek) and Tie-1 endothelial-specific receptor PTKs in vascular development^{25,30-32}. Biochemical data indicate that the rasGAP complex, including p190-rhoGAP and p62, are substrates for Flt-1 and Flk-1 tyrosine-kinase activity^{33,35}. The abnormal endothelial cell-migration patterns observed in *Gap* mutant embryos suggest that rasGAP might be a crucial target of endothelial receptor PTKs.

The idea that endothelial cell phenotypes observed in *Gap* mutants might be due to inappropriate signalling by Ras and related GTPases is supported by *in vitro* studies indicating that Ras is required for the motility of cultured endothelial cells, and that oncogenic Ras variants induce a 'random-walk' trajectory^{36,37}. Ras-GTP can also apparently stimulate a GTPase cascade³⁸ by activating Rac, which is upstream of Rho. In Swiss 3T3 cells the Cdc42, Rac and Rho GTPases regulate the assembly of specific focal complexes associated with filopodia, lamellipodia and actin stress fibres, respectively³⁹. Thus the absence of rasGAP could potentially dysregulate several different GTPases, including Ras (directly), Rac (indirectly through an effect on

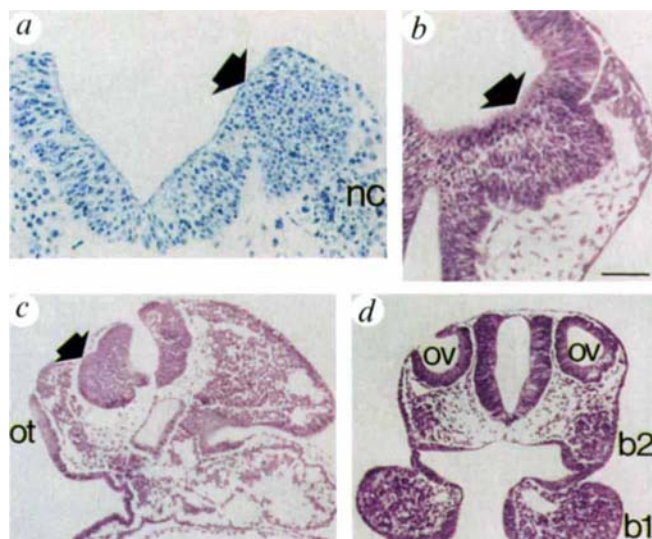


FIG. 5 Excess neuronal cells in *Gap*; *Nf1* double homozygotes. a-c. Transverse sections of three different *Gap*; *Nf1* double mutants detail localized regions of increased neuronal cells in the hindbrain (arrowheads). a, Higher magnification of the section shown in Fig. 4e shows an excess of neuronal cells in rhombomere r4 as indicated by a thickening of the neural tube and the presence of apoptotic neural crest (nc) cells. This double mutant embryo also exhibited another bulge in r1/r2 (not shown). The double homozygote in b exhibited a protruberance in r1/r2 and, in a third specimen, outgrowths were detected in r5/r6 (c) and in the posterior neural tube (not shown). This abnormal accumulation of cells in the neural tube is specific for *Gap*; *Nf1* double homozygotes and has not been observed in other embryos as shown for a *Nf1* single mutant at the level of r5/r6 (d). Note that this single mutant exhibited proper dorsal closure of the hindbrain and formed normal otic vesicles (ov) and branchial arches (b1 and b2). Scale bars: a, b, 50 μ m; c, d, 100 μ m.

Ras), and Rho (through altered regulation of p190-rhoGAP or Rac), with consequent disruption of cell architecture and movement.

An intriguing aspect of the *Gap* mutant phenotype involves the striking increase in apoptotic death of cells that make up the anterior neural tube and cranial neural crest. This suggests that rasGAP might function downstream of PTKs in the nervous system to monitor the cues that control cell proliferation, differentiation, survival or death. By controlling Ras-GTP levels, rasGAP could prevent neuronal cells from undergoing premature entry into the cell cycle, and thereby outstripping their neurotrophic support. This function is consistent with observations⁴⁰ that, without inhibiting Ras signalling, trophic-deprived PC12 cells make a failed attempt to enter the cell cycle and activate a programme of cell death.

Alternatively, rasGAP might regulate specific apoptotic signalling pathways. Both rasGAP and neurofibromin have been shown⁴¹ to function *in vitro* as GAPs for R-Ras, a Ras-related protein that also binds⁴² to Bcl-2, which is implicated in the control of apoptotic signalling pathways. R-Ras, Ha-Ras and Rho have all been shown to promote apoptosis of cultured cells under specific conditions⁴³⁻⁴⁵. The detection of massive neuronal cell death in early-stage *Gap* mutant embryos provides an *in vivo* linkage between Ras signalling and apoptotic death pathways.

It is apparent that *Gap* has biological functions that are not redundant with *Nf1*, as they exhibit distinct recessive phenotypes. However, embryos deficient for both rasGAP and neurofibromin show a more severe phenotype than results from either mutation on its own. The enhanced phenotype of *Gap*; *Nf1* double homozygotes also suggests that Ras proteins participate in additional developmental processes, including dorsal closure of the neural tube, otic vesicle formation, mesoderm/somite patterning, posterior elongation and embryo turning. The genetic

interaction between *Gap* and *Nfl* may reflect increasing Ras deregulation, as two of the four known GAPs specific for Ras proteins are eliminated. Furthermore, the highly localized patches of excess neuronal cells detected in the double homozygotes suggests that excess Ras signalling in mouse embryos can alter specific cell proliferation or survival pathways.

Taken together, these results show that rasGAP (and by extension Ras) is essential for control of cell survival and movement

during embryonic development. The functions of rasGAP differ between different cell types, and are also distinct from those of neurofibromin, although genetic evidence suggests that these two GAPs act in synergy to regulate Ras *in vivo*. The lethal phenotype resulting from the mutation in *Gap* is consistent with the view that widely expressed SH2 signalling proteins are crucial to normal cell functions, probably by coordinating events downstream of tyrosine kinases. □

Received 5 June; accepted 25 September 1995.

1. Barbacid, M. A. *Rev. Biochem.* **56**, 779–827 (1987).
2. Greenwald, I. & Rubin, G. M. *Cell* **68**, 271–281 (1992).
3. McCormick, F. *Curr. Opin. Genet. Dev.* **5**, 51–55 (1995).
4. Boguski, M. S. & McCormick, F. *Nature* **366**, 643–654 (1993).
5. Rodriguez-Viciana, P. et al. *Nature* **370**, 527–532 (1994).
6. Trahey, M. et al. *Science* **242**, 1697–1700 (1988).
7. Vogel, U. S. et al. *Nature* **335**, 90–93 (1988).
8. Ballester, R. et al. *Cell* **63**, 851–859 (1990).
9. Martin, G. A. et al. *Cell* **63**, 843–849 (1990).
10. Xu, G. F. et al. *Cell* **63**, 835–841 (1990).
11. Maekawa, M. et al. *Molec. cell. Biol.* **14**, 6879–6885 (1994).
12. Cullen, P. J. *Nature* **376**, 527–530 (1995).
13. Pawson, T. *Nature* **373**, 573–580 (1995).
14. Molloy, C. J. et al. *Nature* **342**, 711–714 (1989).
15. Anderson, D. et al. *Science* **250**, 979–982 (1990).
16. Ellis, C., Moran, M., McCormick, F. & Pawson, T. *Nature* **343**, 377–381 (1990).
17. Kaplan, D. R., Morrison, D. K., Wong, G., McCormick, F. & Williams, L. T. *Cell* **61**, 125–133 (1990).
18. Kazlauskas, A., Ellis, C., Pawson, T. & Cooper, J. A. *Science* **247**, 1578–1581 (1990).
19. Moran, M. F., Polak, P., McCormick, F., Pawson, T. & Ellis, C. *Molec. cell. Biol.* **11**, 1804–1812 (1991).
20. Settleman, J., Narasimhan, V., Foster, L. C. & Weinberg, R. A. *Cell* **69**, 539–549 (1992).
21. Ridley, A. J. & Hall, A. *Cell* **70**, 389–399 (1992).
22. Marshall, C. J. *Trends Genet.* **7**, (1991).
23. Gasca, S., Hill, D. P., King-Smith, J. & Rossant, J. *Dev. Genet.* **17**, 141–154 (1995).
24. Brannan, C. I. et al. *Genes Dev.* **8**, 1019–1029 (1994).
25. Conlon, R. A., Reaume, A. G. & Rossant, J. *Development* **121**, 1533–1545 (1995).
26. Fong, G.-H., Rossant, J., Gertszenstein, M. & Breitman, M. L. *Nature* **376**, 66–70 (1995).
27. Shalaby, F. et al. *Nature* **376**, 62–66 (1995).

28. Sato, T. N. et al. *Nature* **376**, 70–74 (1995).
29. Waitenberger, J., Claesson-Welsh, L., Sjöberg, A., Shibuya, M. & Heidén, C.-H. *J. Biol. Chem.* **269**, 26988–26995 (1994).
30. Seetharam, L. et al. *Oncogene* **10**, 135–147 (1995).
31. Gou, D., Jia, Q., Song, H.-Y., Warren, R. S. & Donner, D. B. *J. Biol. Chem.* **270**, 6729–6733 (1995).
32. Sosnowski, R. G., Feldman, S. & Feramisco, J. R. *J. Cell Biol.* **121**, 113–119 (1993).
33. Fox, P. L., Sa, G., Dobrowolski, S. F. & Stacey, D. W. *Oncogene* **9**, 3519–3526 (1994).
34. Nobes, C. D., Hawkins, P., Stephens, L. & Hall, A. *J. Cell Sci.* **108**, 225–233 (1995).
35. Nobes, C. D. & Hall, A. *Cell* **81**, 53–62 (1995).
36. Ferrari, G. & Green, L. A. *EMBO J.* **13**, 5922–5928 (1994).
37. Rey, I., Taylor-Harris, P., van Erp, H. & Hall, A. *Oncogene* **9**, 685–692 (1994).
38. Fernandez-Sarabia, M. J. & Bischoff, J. R. *Nature* **366**, 274–275 (1993).
39. Jimenez, B. et al. *Oncogene* **10**, 811–816 (1995).
40. Tanaka, N. et al. *Cell* **77**, 829–839 (1994).
41. Wang, H.-G. et al. *J. Cell Biol.* **129**, 1103–1114 (1995).
42. Tybulewicz, V. L. J., Crawford, C. E., Jackson, P. K., Bronson, R. T. & Mulligan, R. C. *Cell* **65**, 1153–1163 (1991).
43. Hsieh, C.-L., Vogel, U. S., Dixon, R. A. F. & Francke, U. *Somat. Cell molec. Genet.* **15**, 579–590 (1989).
44. Wurst, W. & Joyner, A. L. in *Gene Targeting: A Practical Approach* (ed. Joyner, A. L.) 31–62 (Oxford Univ. Press, 1993).
45. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. *Proc. natn. Acad. Sci. U.S.A.* **90**, 8424–8428 (1993).
46. Kim, H. A. et al. *Oncogene* **11**, 325–335 (1995).

ACKNOWLEDGEMENTS. We thank M. Moran for the complementary DNA clone encoding murine Gap; A. Reaume for the 129 strain genomic library; G. Martin for anti-rasGAP antibodies; C. Elliott and S. Kulkarni for technical assistance; J. Rossant for generating the *Gap* germline transmitting chimaeras and for encouragement; S. Gasca for the *Cordon-bleu* LacZ marker; A. Wyllie for help in the analysis of the electron microscopy data; and T. Saxton, R. Dhand and N. Al-Alawi for critical reading of the manuscript. This work was supported by grants from Bristol-Myers-Squibb, a Terry Fox programme grant from the National Cancer Institute of Canada (NCIC), an International Research Scholar Award from the Howard Hughes Medical Institute (HHMI) to T.P., T.J. is an Assistant Investigator of the HHMI, and T.P. is a Terry Fox Cancer Research Scientist of the NCIC. M.H. was supported by a postdoctoral fellowship from the Medical Research Council of Canada.

Near-coeval formation of the Galactic bulge and halo inferred from globular cluster ages

Sergio Ortolani*, Alvio Renzini†‡, Roberto Gilmozzi‡, Gianni Marconi§, Beatriz Barbuy||, Eduardo Bica¶ & R. Michael Rich*

* Università di Padova, Dipartimento di Astronomia, Padova, Italy

† Università di Bologna, Dipartimento di Astronomia, Bologna, Italy

‡ European Southern Observatory, Garching, Germany

§ Osservatorio Astronomico di Roma, Monte Porzio, Roma, Italy

|| Universidade de São Paulo, Departamento de Astronomia, São Paulo, Brazil

¶ Universidade Federal do Rio Grande do Sul, Departamento de Astronomia, Porto Alegre, Brazil

Department of Astronomy, Columbia University, New York, USA

THE morphology of our Galaxy is characterized by a disk of stars moving on circular orbits, surrounding a central spheroidal body of stars on high-velocity, randomly oriented orbits. The spheroid is further differentiated into an inner bulge and an outer halo; the bulge stars are rich in elements heavier than helium ('metals'), whereas the halo stars are metal-poor, suggesting that the latter formed very early in the history of the Galaxy. (They

have experienced little chemical enrichment, by previous generations of stars.) It is not known, however, whether the bulge is the inner extension of the halo, having formed as part of the same process¹, or whether it formed much later, perhaps by a dynamical distortion of the inner regions of the disk^{2,3}. Here we report observations obtained with the Hubble Space Telescope of two metal-rich globular clusters that form part of the bulge population. Within the uncertainties, these bulge globular clusters appear to be coeval with halo clusters, which suggests that the formation of the bulge was part of the dynamical process that formed the halo, and that the bulge gas underwent rapid chemical enrichment, in less than a few billion years.

There are two versions of the option in which halo and bulge share a common origin. In one view, star formation starts out in the halo, and gas progressively metal-enriched by supernova explosions drips towards the centre, so that the innermost part of the spheroid—the bulge—forms last as the culmination of the condensation process⁴. In the other view the bulge forms first, as star formation first erupts in the central, high-density regions of the protogalaxy, rapidly enriching the gas to solar metallicity and beyond, until local heating quenches further star formation^{5–8}. A wide distribution of stellar ages, from ~15 Gyr to just a few Gyr, is instead expected if the bulge is formed late, a result of the dynamical evolution of the inner disk. Therefore, a decisive step towards distinguishing between all these competing scenarios must come from determining the age of stars in the bulge.

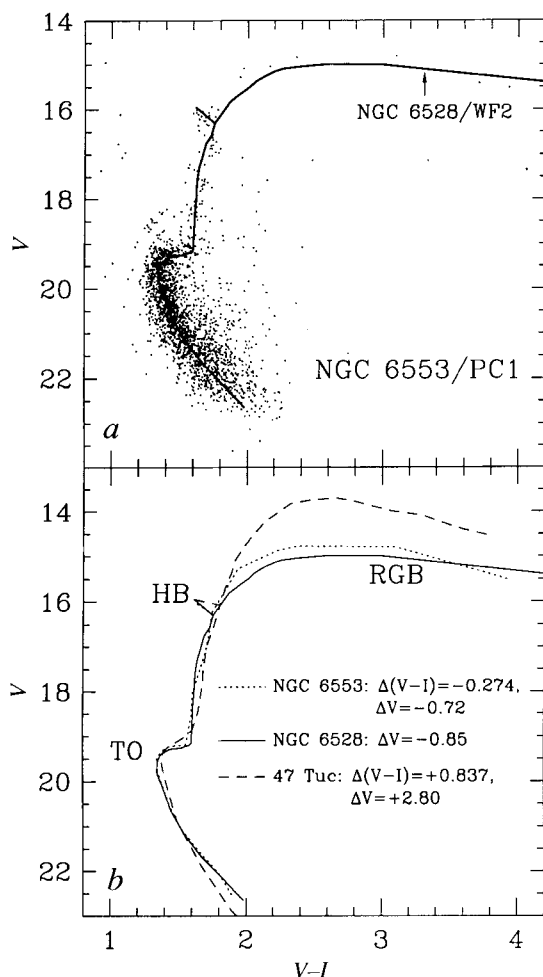


FIG. 1 *a*, The colour-magnitude diagram (CMD) of the cluster NGC6553 for the Planetary Camera (WFPC2-PC1) data. The HST filters F555W and F814W have been used for *V* and *I*, with undithered exposures of 2×100 and 2×50 s, respectively in *V* and *I*. Shallow exposures (14 s in *V* and 5 s in *I*) ensured accurate photometry for the bright stars. Standard STScI reductions and calibrations and the DAOPHOT²⁷ photometric package have been used (the latter with point spread function fitting). Note the stubby red horizontal branch (HB) at $V \approx 16.2$. The modest bright extension of the main sequence is due to contamination of disk stars in the spiral arm at 2 kpc (ref. 14), blending of stellar images due to crowding, and (possibly) blue stragglers. Superimposed is the cluster mean locus obtained from the WFC2 data of NGC6528. Note how well the mean locus of NGC6528 fits the NGC6553 data. The cluster mean locus represents the ridge-line of the star distribution, and operationally is obtained by constructing colour distributions for each luminosity bin and then taking the median colour. The data points of NGC6553 have been de-reddened as indicated in *b*, so to make its turnoff colour equal to that of NGC6528. The locus of NGC6528 has been shifted by $\Delta V = -0.85$, to make it match the CMD of NGC6553. In this way the two CMDs have been reduced to the same, lowest reddening and distance (respectively that of NGC6528 and NGC6553). Thus the true distance modulus of NGC6553 is almost 1 mag smaller than that of NGC6528. Hence, if NGC6528 is close to the Galactic Centre (that is, at ~ 8 kpc from us), NGC6553 is at a distance of ~ 5 kpc from us, ~ 3 kpc from the Galactic Centre, at the outer fringe of the bulge. *b*, The mean cluster loci for NGC6528 (same as in *a*) and NGC6553 (from PC1), full and dotted line, respectively. The two loci have been shifted as described above, and indicated in the figure. Also shown is the mean locus of the inner halo cluster 47 Tuc (dashed line), for which $[\text{Fe}/\text{H}] \approx -0.7$. It has been shifted to bring into coincidence the blue end of its HB with that of the bulge clusters for the sole purpose of emphasizing the very close similarity in HB to turnoff luminosity difference (the age indicator), and the brighter upper red-giant branch in 47 Tuc, indicating that this cluster is much less metal-rich than the two bulge clusters. Note also the steeper subgiant branch of 47 Tuc (from the TO to the base of the red-giant branch (RGB)), also indicative of a lower metallicity compared to the two bulge clusters.

Recent stellar-population studies of bulge fields all agree that the bulk of stars in the bulge are very old, older than at least 5–8 Gyr, but diverge on whether the data require the presence of a population substantially younger than 15 Gyr (refs 8–15). Precise age dating of bulge stars is indeed complicated by several factors, including crowding, depth effects, variable obscuration (reddening), metallicity dispersion, and contamination by foreground disk stars, all effects that can mimic or mask intrinsic age variations. Many of these limitations can be reduced by studying globular star clusters that are physically located in the bulge, and whose high metallicity ensures that they are not halo interlopers. Indeed, globular clusters are compact, populous groups of stars all of the same age and composition, and we have selected the clusters NGC6528 and NGC6553 for study by the Hubble Space Telescope (HST). These clusters are respectively located at $\sim 4^\circ$ and 6° from the Galactic Centre, and their metallicity is about solar, that is, $[\text{Fe}/\text{H}] \equiv \log (\text{Fe}/\text{H})_{\text{cluster}} - \log (\text{Fe}/\text{H})_{\odot} = 0.0 \pm 0.3$ (refs 16–18), close to the average for stars in the bulge¹⁹. Their radial velocities and distances are also consistent with membership of the bulge^{20,21}.

From our HST observations, using the Wide Field Planetary Camera (WFPC2), accurate colour-magnitude diagrams have been obtained for each cluster and for each of the four CCD chips of the camera that sample adjacent portions of each cluster. Figure 1*a* shows the colour-magnitude diagram of NGC6553 as sampled by the Planetary Camera chip (PC1); the smaller pixel size allows a better sampling of stellar images, resulting in higher photometric accuracy compared to the Wide Field Camera chips. The dispersion of the diagram is actually dominated by differential reddening (~ 0.2 mag across the field) due to variable amounts of interstellar absorption along the line of sight, rather than by photometric errors ($\lesssim 0.1$ mag for $V \lesssim 20$). However, mean loci—defined by the ridge-line of the distribution of the stars in the colour-magnitude diagram—were constructed for each of the four chips. They appear identical apart from a chip-to-chip shift due to the different mean reddening in each chip.

The same procedure has been repeated for the NGC6528 data, and in Fig. 1*b* we compare the cluster locus of NGC6528 (WFC2 chip) to that of NGC6553 (PC1). The two loci are remarkably similar, from the lower main-sequence all the way to the tip of the red-giant branch, as further illustrated in Fig. 1*a* where the WFC2 locus of NGC6528 is superimposed on the PC1 data for NGC6553. In particular, the luminosity differences $\Delta V_{\text{TO}}^{\text{HB}}$ between the horizontal branch (HB) and the main-sequence turnoff (TO) are the same to within better than ~ 0.1 mag. As apparent in both panels, $\Delta V_{\text{TO}}^{\text{HB}}$ may perhaps be ~ 0.05 mag smaller in NGC6553. We do not consider this difference to be

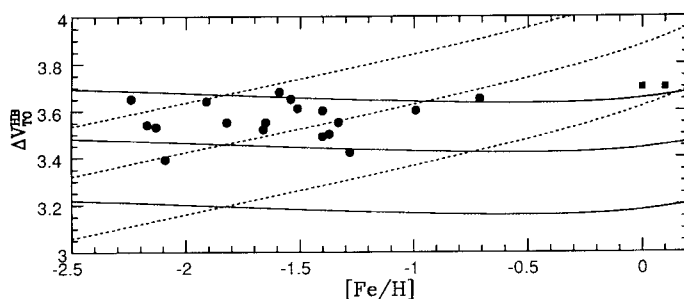


FIG. 2 The luminosity difference between the HB and the turnoff $\Delta V_{\text{TO}}^{\text{HB}}$ is plotted for a representative set of halo globular clusters²² (filled circles) and for NGC6528 and NGC6553 (filled squares). For this display $[\text{Fe}/\text{H}] = 0.0$ and 0.1 has been adopted for these two clusters. The lines refer to ages of 18, 15 and 12 Gyr (upper, middle and lower lines, respectively), with these assumptions: (1) turnoff luminosity-age relation from ref. 22; (2) helium abundance $Y = 0.23 + 3.0 \times Z$; (3) no enhancement of α -elements (that is, solar proportions); (4) HB luminosity-metallicity relations $M_V^{\text{HB}} = 1.17 + 0.39[\text{Fe}/\text{H}]$ (ref. 24, solid lines); and (5) $M_V^{\text{HB}} = 0.73 + 0.15[\text{Fe}/\text{H}]$ (ref. 25, dotted lines).

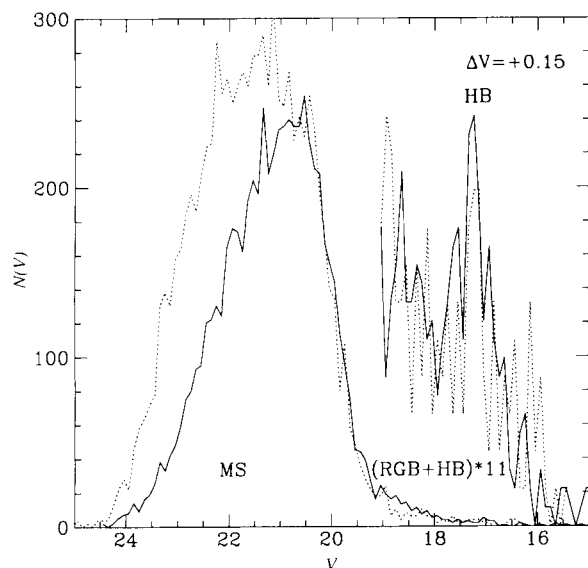


FIG. 3 The luminosity function of main-sequence (MS) and red-giant (RGB+HB) stars in NGC6528 (WFC2 field, dotted line) and in Baade's Window (solid lines), the bulge low-reddening region with ~ 500 pc projected distance from the Galactic Centre. Main-sequence stars are defined as those fainter than $V=10(V-I)+2.65$ in Baade's Window, and $V=10(V-I)+4.5$ in NGC6528. The luminosity function of the cluster has been shifted by $\Delta V = +0.15$ so as to bring into coincidence its HB peak (marked on the figure) with that of Baade's Window, and multiplied by a factor of 2 so as to normalize the two distributions at $V=20.45$, where both are reasonably complete. An identical multiplying factor is obtained when normalizing the two distributions to the same number of RGB+HB stars brighter than $V=19.45$ in this figure. Note that below $V \approx 21$ the luminosity function of the bulge is progressively more incomplete compared to that of the cluster. The luminosity function of the cluster has been broadened with a Monte Carlo simulation to mimic the depth effect present in the Baade's Window field, by adopting a $\rho \propto r^{-3.7}$ density law for the bulge²⁸. For this figure, the RGB+HB luminosity functions have been multiplied by a factor of 11, to avoid overlap with the luminosity function of the disk stars in the spiral arm at 2 kpc (ref. 14). The Baade's Window data have been obtained at the ESO/NTT telescope with ~ 0.4 -arcsec seeing (S.O., A.R. and R.M.R., manuscript in preparation). The crowding conditions tested by artificial star experiments are quite similar to that of the cluster WFC2 data, and therefore crowding-related photometric errors have quite similar effects on the two luminosity functions here displayed, for the part where they are virtually complete.

significant, given the uncertainties on the mean loci induced by crowding-related photometric errors and by the limited number of stars defining the horizontal branch. Seemingly the same are the colour differences between the main-sequence turnoff and the base of the red-giant branch. Both luminosity and colour of the turnoff are sensitive functions of cluster age. Specifically, $\delta \Delta V_{\text{TO}}^{\text{HB}} \approx \delta \ln(\text{age})$ (ref. 22), which ensures that the two clusters have the same age to within $\sim 10\%$. The identity of the upper red-giant branches further ensures that the two clusters have nearly identical metallicity. The maximum V luminosity (above the horizontal-branch level) is the same within ~ 0.2 mag. Because increasing the metal abundance Z by 50% has the effect of decreasing by ~ 0.4 mag this maximum V luminosity²³, the metallicity difference between the two clusters is $\lesssim 20\%$.

Having established their identity, we now consider the age of the two clone clusters. This is less easy to determine, as subtle ambiguities remain in both the observations and the theory. Cluster absolute ages are best inferred from a semi-empirical calibration of the age- $\Delta V_{\text{TO}}^{\text{HB}}$ relation²². For halo globulars, the horizontal-branch luminosity is measured at the turnoff colour, whereas the horizontal branches of our clusters do not extend that far bluewards, and the luminosity of these branches increases towards the blue, an effect due to a combination of differential reddening and higher TiO blanketing towards the red. Thus, $\Delta V_{\text{TO}}^{\text{HB}} = 3.85$ or 3.5 , depending on which of the two extremes of the horizontal-branch locus in Fig. 1 is used. We adopt $\Delta V_{\text{TO}}^{\text{HB}} = 3.7 \pm 0.15$. For comparison, Fig. 1 also shows the mean locus of 47 Tuc (S.O., A.R. and R.M.R., manuscript in preparation), one of the most metal-rich clusters in the inner halo (but still ~ 5 times more metal poor than the two bulge clusters). Clearly its $\Delta V_{\text{TO}}^{\text{HB}}$ is less than or equal to that of the two bulge clusters. This ensures that the age of the three clusters is nearly the same, thus establishing that the two bulge clusters are nearly coeval with the metal-rich halo clusters. Next comes the problem of the calibration. For illustration we use two extreme calibrations of the $M_V^{\text{HB}} - [\text{Fe}/\text{H}]$ relation^{24,25}, and the result is shown in Fig. 2. Not surprisingly, depending on which relation is used the bulge clusters appear either slightly older than the metal-poor clusters of the halo, or some billions of years younger than them, in the latter case continuing an apparent age-metallicity trend^{25,26}. Compelling evidence in favour of one calibration or the other is still lacking, but soon HST observations of globular clusters in the Andromeda galaxy may decide the issue. Absolute age dating also requires measurement of α -

element abundances, a project that we are pursuing. Notwithstanding these limitations, we emphasize that the continuity of the $\Delta V_{\text{TO}}^{\text{HB}} - [\text{Fe}/\text{H}]$ relation from the halo all the way to the bulge is sufficient to support the contention that the whole spheroid was formed in a single, continuous, rapid process, rather than the bulge having been added later by a different process. Also worth noting is that the high metallicity of the two clusters was reached in a rather short time (less than a few billion years) over the whole bulge, up to ~ 3 kpc from the Galactic Centre. We interpret this evidence as favouring a fast, dissipative formation of the bulge, though we are not yet able to distinguish between the inside-out and outside-in scenarios.

We now consider if our two clusters are representative of the general population of the bulge. We do this by comparing the stellar population of the clusters with that of the bulge field known as Baade's Window, which combines relatively low reddening and high average metallicity¹⁹. This comparison reveals that the age and metallicity of the bulge population in the Window must be very similar to that of the two clusters. This is illustrated in Fig. 3, showing the luminosity functions (the number of stars per magnitude bin) of NGC6528 (WFC2) and Baade's Window. When the horizontal-branch peaks in the luminosity function are made to coincide, the main-sequence luminosity functions also coincide with extremely high precision in the brighter, age-dependent part that is less affected by incompleteness and field contamination (that is, $19.5 \lesssim V \lesssim 20.5$). Given that the relative positioning of the two luminosity functions is accurate to within ~ 0.1 mag, we infer that the age of the bulk population in Baade's Window is the same as that of the two globulars NGC6528 and NGC6553 to within 10% . The identical slope of the main-sequence luminosity functions also ensures that the age dispersion of Baade's Window stars must also be very small, probably $\lesssim 10\%$. We conclude that the cluster and Baade's Window data strongly support the notion of a fast formation and chemical enrichment of the bulge as part of the spheroid formation process, early in the evolution of the Galaxy. $\square$

Received 3 April; accepted 13 September 1995.

1. Eggen, O. J., Lynden-Bell, D. & Sandage, A. *Astrophys. J.* **136**, 748–766 (1962).
2. Raha, N., Sellwood, J. A., James, R. A. & Kahn, F. D. *Nature* **352**, 411–412 (1991).
3. Pfenniger, D. & Friedli, D. *Astr. Astrophys.* **252**, 75–93 (1991).
4. Larson, R. B. *Mon. Not. R. Astr. Soc.* **173**, 671–699 (1975).
5. Renzini, A. & Greggio, L. in *Bulges of Galaxies 47–63* (eds Jarvis, B. J. & Terndrup, D. M.) (Conf. & Workshop Ser. 35, ESO, Garching, 1990).

6. Larson, R. B. *Publ. astr. Soc. Pacif.* **102**, 709–722 (1990).
7. Lee, Y. W. *Astr. J.* **104**, 1780–1789 (1992).
8. Renzini, A. in *Galactic Bulges* 151–168 (eds Dejonghe, H. & Habing, H. J.) (IAU Symp. 153, Kluwer, Dordrecht, 1993).
9. Rich, R. M. *Mem. Soc. Astr. It.* **56**, 23–35 (1985).
10. Frogel, J. A. *Rev. Astr. Astrophys.* **26**, 51–92 (1988).
11. Terndrup, D. M. *Astr. J.* **96**, 884–908 (1988).
12. Whitelock, P., Feast, M. & Catchpole, R. *Mon. Not. R. astr. Soc.* **248**, 276–312 (1991).
13. Holzman, J. A. *et al. Astr. J.* **106**, 1826–1838 (1993).
14. Paczyński, B. *et al. Astr. J.* **107**, 2060–2066 (1994).
15. Renzini, A. *Astr. Astrophys.* **285**, L5–L8 (1994).
16. Zinn, R. *Astrophys. J.* **293**, 424–444 (1985).
17. Barbuy, B., Castro, S., Ortolani, S. & Bica, E. *Astr. Astrophys.* **259**, 607–613 (1992).
18. Cohen, J. G. & Sleeper, C. *Astr. J.* **109**, 242–263 (1995).
19. McWilliam, A. & Rich, R. M. *Astrophys. J. Suppl. Ser.* **91**, 749–791 (1994).
20. Hesser, J. E., Shawl, S. J. & Meyer, J. E. *Publ. Astr. Soc. Pacif.* **98**, 403–422 (1986).
21. Minniti, D. *Astr. J.* **109**, 1663–1669 (1995).
22. Renzini, A. in *Observational Tests of Cosmological Inflation* 131–146 (eds Shanks, T. *et al.*) (Kluwer, Dordrecht, 1991).
23. Garnavich, P. M., VandenBerg, D. A., Zurek, D. R. & Hesser, J. E. *Astr. J.* **107**, 1097–1110 (1994).
24. Sandage, A. & Cacciani, C. *Astrophys. J.* **350**, 645–661 (1990).
25. Walker, A. R. *Astrophys. J.* **390**, L81–L84 (1992).
26. Carney, B. *Mem. Soc. Astr. It.* **63**, 409–431 (1992).
27. Stetson, P. *Publ. Astr. Soc. Pacif.* **99**, 191–222 (1987).
28. Sellwood, J. & Sanders, R. *Mon. Not. R. astr. Soc.* **233**, 611–620 (1988).

ACKNOWLEDGEMENTS. This work is based on observations with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, operated by AURA Inc. under contract to NASA, and with the European Southern Observatory NTT telescope at La Silla, Chile. R.M.R. is an A. P. Sloan Fellow at Columbia University. This work was supported in part by ASI, and by NASA (R.M.R.).

Formation of fold-and-thrust belts on Venus by thick-skinned deformation

M. T. Zuber*† & E. M. Parmentier‡

* Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, Maryland 21218, USA

† Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge Massachusetts 02139-4307, USA

‡ Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA

ON Venus, fold-and-thrust belts—which accommodate large-scale horizontal crustal convergence—are often located at the margins of kilometre-high plateaux^{1–5}. Such mountain belts, typically hundreds of kilometres long and tens to hundreds of kilometres wide, surround the Lakshmi Planum plateau in the Ishtar Terra highland (Fig. 1). In explaining the origin of fold-and-thrust belts, it is important to understand the relative importance of thick-skinned deformation of the whole lithosphere and thin-skinned, large-scale overthrusting of near-surface layers. Previous quantitative analyses of mountain belts on Venus have been restricted to thin-skinned models^{6–8}, but this style of deformation does not account for the pronounced topographic highs at the plateau edge. We propose that the long-wavelength topography of these venusian fold-and-thrust belts is more readily explained by horizontal shortening of a laterally heterogeneous lithosphere. In this thick-skinned model, deformation within the mechanically strong outer layer of Venus controls mountain building. Our results suggest that lateral variations in either the thermal or mechanical structure of the interior provide a mechanism for focusing deformation due to convergent, global-scale forces on Venus.

Distinguishing between thin- and thick-skinned styles of deformation was a long-running controversy in understanding the origin of terrestrial mountain belts, and both styles of deformation are now understood to play a role⁹, sometimes in the same mountain belt^{10–13}. Classic examples of thin-skinned fold-and-thrust belts on Earth include Taiwan¹⁴ and the Canadian Rockies⁹; thick-skinned examples include the Grenville front¹⁵ and the Pyrenees¹⁶. The best-developed fringing mountain belts

on Venus occur in the Ishtar Terra highland (Fig. 1), though crustal plateaux such as Alpha and Tellus Regiones also exhibit the highest elevations at their margins¹⁷. Although previous models favoured a thin-skinned explanation for these features^{6–8}, structural mapping from Magellan images and other observations^{18–21} have prompted the qualitative suggestion²² of deeper deformation contributing to mountain-belt morphology.

Here we present a first quantitative test of the thick-skinned deformation hypothesis. Specifically, we explore the possibility that broad-scale mountain-belt/plateau morphology can be explained by horizontal shortening of a lithosphere that is laterally heterogeneous, due to a change in thickness of either the thermal lithosphere or the crust. Lateral heterogeneities in litho-

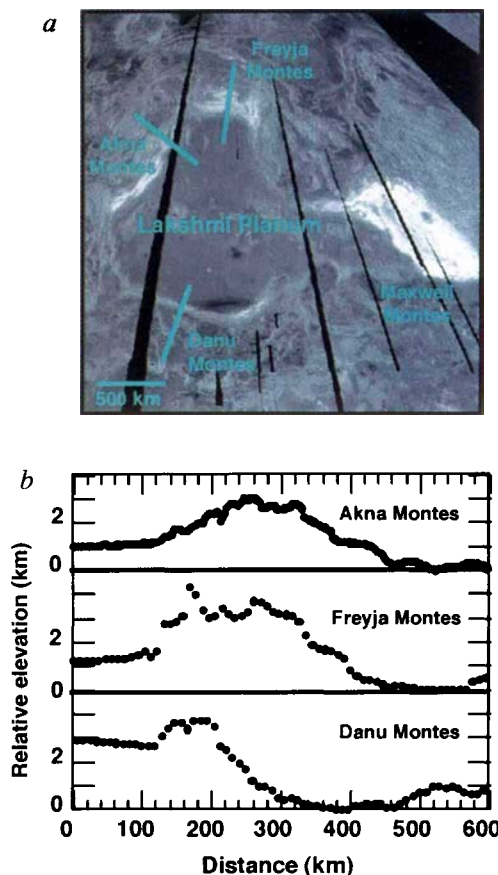


FIG. 1 Fold-and-thrust belts at the edges of the Lakshmi Planum plateau in the Ishtar Terra highland of Venus. *a*, Magellan synthetic aperture radar mosaic (C2-MIDR 60N333;2) showing profile locations. The mountain belts appear as narrow zones of high radar backscatter and are brighter than the central, smooth (radar-dark) Lakshmi Planum plateau. The very bright region to the far right (east) is Maxwell Montes, the highest area of Venus. The black spaces are gaps in the synthetic aperture radar coverage. *b*, Magellan topographic profiles⁴⁰. In all profiles the Lakshmi Planum plateau is on the left. The mountain belts at the edge of the plateau are a characteristic feature of fold-and-thrust belt morphology on Venus. From top to bottom: Akna Montes (long. 324.7°, lat. 68.5°), Freyja Montes (long. 329.8°, lat. 70.1°), and Danu Montes (long. 320.2°, lat. 55.7°), where locations correspond to that at the plateau end of the profiles. In the Akna profile there is a 1-km elevation difference between the tessera terrain at lower relative elevation and plateau, and the Akna mountains rise 2 km above the plateau. In the Freyja Montes profile there is a 1-km elevation difference between the plateau and surroundings and the mountains are elevated by 3 km above the plateau. In the Danu Montes profile there is a 3-km elevation difference between the plateau and surroundings and the mountains rise 1 km above the plateau. The lowland trough and outer rise on the far right of the Danu profile may represent evidence for flexure of the Venus lithosphere in response to the load of the highlands. Vertical exaggeration is 32:1.

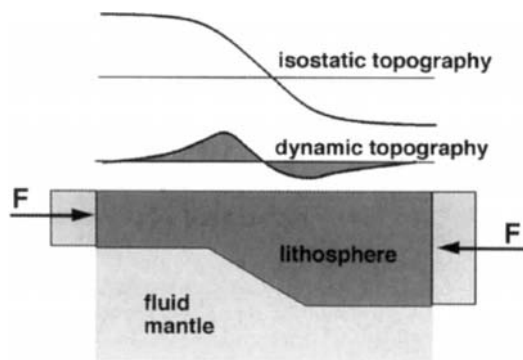


FIG. 2 Horizontal shortening of a lithosphere of variable thickness creates dynamic stress-supported surface topography. The bottom part of the figure shows that horizontal forces (F), representing the resultant of stress distributions within the lithosphere, have equal magnitude but different lines of action due to changes in lithosphere thickness. The moment thus created must be balanced by an opposing moment due to gravity that acts on dynamic surface topography. With no applied vertical forces, the average topography and the net gravitational attraction must vanish. This stress-supported topography, illustrated in the central part of the figure, combined with isostatic topography, illustrated in the top part of the figure, provides a 'thick-skinned' mechanism to explain the observed morphology of venusian mountain belts. The profile illustrates the deformation pattern that would develop in a thin, uniform viscosity layer without buoyancy forces and assuming infinitesimal strain⁴¹.

sphere structure can develop in response to thermal thinning by the delamination or gravitational instability of the deeper weaker lithosphere, whereas variations in crustal thickness may arise due to either spatially variable melting of the mantle or by horizontal shortening of the crust. Both shallow crustal and deeper thermal-density variations may be required to explain gravity data in Ishtar Terra^{22,23}.

In a variable-thickness lithosphere or crust that is shortened horizontally, dynamic moments will produce stress-supported topography. This mechanism, illustrated schematically in Fig. 2, may provide a simple conceptual explanation for the topographic high at the edge of a plateau. In this model, isostatic effects due to subsurface compositional and thermal variations also contribute to the shape and elevation of the plateau. Flexure associated with thin-skinned deformation may be an alternative

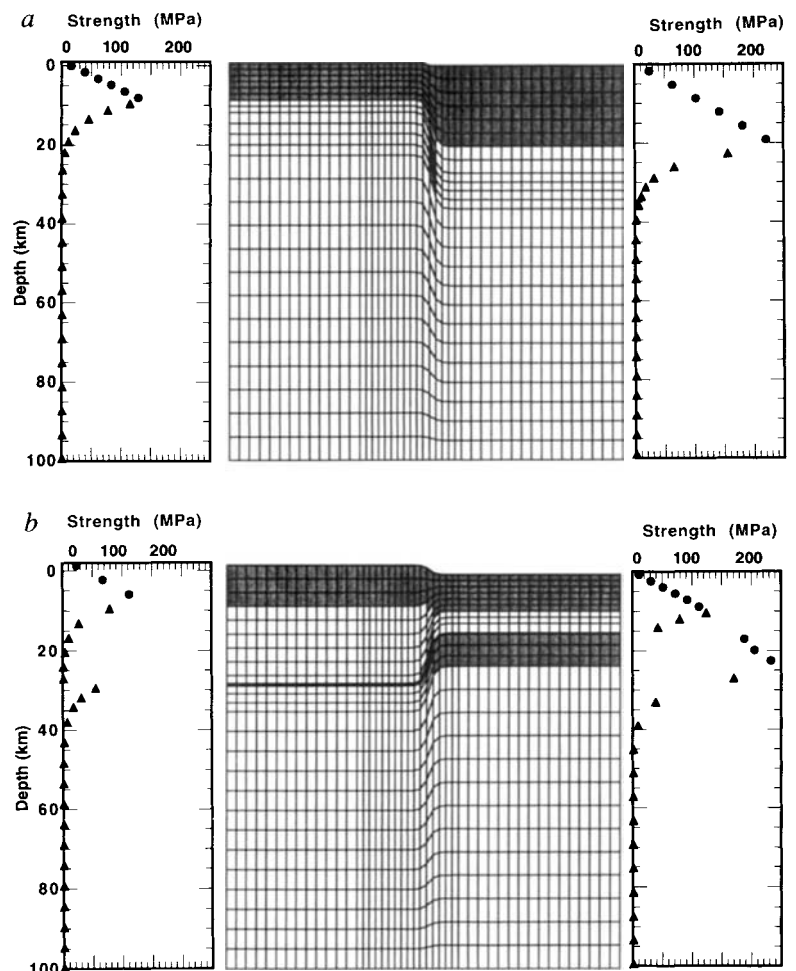
explanation for the topographic lows sometimes observed in the surroundings⁸, but flexural models offer no simple explanation for the topographic high, which is a much more distinct and persistent feature than a surrounding low.

As this model suggests, the style of deformation is sensitive to the rheological structure. To formulate more realistic models that consider both horizontal and vertical variations in mechanical properties, we employ finite-element solutions based on a standard penalty function formulation for incompressible viscous flow^{24,25}. In the non-newtonian viscous lithosphere we adopt a strain-rate dependent viscosity, $\mu(x, z, t)$, of the form²⁵

$$\mu = \mu_0 [1/\dot{\epsilon}_2]^{(1-1/n)}$$

where μ_0 is the pressure and temperature-dependent part of the viscosity which is advected with each element as the model

FIG. 3 Initial (strain $\epsilon = 0$) finite-element grids and vertical strength profiles for two-dimensional models with a variable-thickness lithosphere (a) and variable-thickness crust (b). The shaded areas of the grid and the filled circles in the strength profiles correspond to zones of brittle ($n = 20$) deformation, where n is an empirical exponent in the stress/strain-rate relationship. The white areas of the grid and the filled triangles represent areas of ductile flow ($n \approx 3$). The model in a assumes an olivine rheology²⁶ and that in b assumes a diabase crust³² and olivine mantle²⁶. The heavy line in b is the crust-mantle boundary. The vertical profiles correspond to initial strengths for the left-most and right-most columns in the grid. Boundary conditions for the models include a top surface which is stress-free and a bottom surface on which both the shear stress and vertical velocity vanish. The right boundary shortens with a constant horizontal velocity, and the left boundary is a symmetry axis. A mean horizontal strain of 10^{-15} s^{-1} , which may be appropriate for deformed venusian terrains⁴², is assumed for both models. A surface thermal gradient of 18 K km^{-1} (refs 17, 31) was assumed for the thicker part of the lithosphere in a and the full lateral extent of the lithosphere in b. Also assumed are a surface temperature of 700 K and a conductive cooling profile. The crustal density is $2,700 \text{ kg m}^{-3}$ and the mantle density is $3,300 \text{ kg m}^{-3}$.



evolves. This assumes that mountain belts form quickly enough that conductive heat transfer does not significantly change temperature of each element and that the overburden pressure of each element remains constant. The latter assumption is not strictly true, but it is a good approximation and simplifies the calculations. The parameter $\dot{\epsilon}_2$ is the second invariant of the strain rate tensor, where $\dot{\epsilon}_2 = (\dot{\epsilon}_{ij}\dot{\epsilon}_{ij})^{1/2}$, $\dot{\epsilon}_{ij}$ is the strain-rate tensor, and n is the power-law exponent of stress in the stress/strain-rate relationship. To simulate the rheology of the lithosphere in the ductile creep regime we used a power-law exponent of ~ 3 ; in the brittle regime we invoked the assumption of perfect plasticity in which n approaches infinity to approximate a material that deforms by pervasive faulting. In the brittle regime, the viscosity was constrained to have an upper bound defined by the plastic yield strength²⁶. The non-newtonian flow was determined using an iterative procedure²⁷, in which the viscosity was recalculated at each time step from the strain rates derived from the velocity field²⁵. Given that the mechanism for the formation of venusian mountain belts and the origin of the convergent forces are matters of debate^{19,23,28–30}, our model assumes uniform horizontal compression without regard to a specific mechanism.

As shown in Fig. 3a, we examine a brittle layer of thickness h initially containing a thickness perturbation of amplitude Δh and half-width L . We adopted a vertical temperature distribution in the lithosphere that corresponds to conductive cooling of a half-space. For both models we incorporated a strength envelope distribution of viscosity with depth³¹, as indicated by

rock-mechanics experiments³². The initial topography is isostatically compensated, and horizontal shortening is imposed by vertical boundaries that converge at a prescribed constant horizontal velocity u . The magnitude of buoyancy forces in resisting the development of surface topography associated with shortening is controlled by the dimensionless parameter $S = \rho g L^2 / \mu_0 u$ where ρ is the crustal density.

Figure 4a shows topographic profiles for horizontal shortening of a variable-thickness lithosphere, with $\Delta h = 0.5 h$ for rheological parameters relevant to Venus³² and a broad range of allowable strains indicated by structural analyses of highlands within Ishtar Terra³³. The model predicts the development of a flanking high at the edge of the topographic rise. For example, the calculated topography can explain the long-wavelength shape of the Akna Montes profile in Fig. 1, both in terms of the change in elevation between the topographic rise and adjacent, lower, tessera terrain and the width and height of the mountain belt at the edge of the plateau, with a thermal gradient of 18 K km^{-1} , a convergence rate of 10^{-15} s^{-1} , and a strain of 0.1. In this model $\sim 80\%$ of the elevation of the mountain belt is supported by stresses in the lithosphere. Our results indicate that the Akna Montes region can be described by the assumed thermal gradient and an age of somewhat greater than 3 Myr, if the rheology we have used is representative of Venus at the time of the deformation and if the other assumptions in our model are correct. The flanking topographic low suggested by the conceptual model in Fig. 2 must be present to achieve a local balance of vertical forces generated by the topography. In the finite-element models, the low is much less pronounced than the topographic high because thicker lithosphere beneath the low makes it much broader than the high and reduces its amplitude.

Corresponding topographic profiles for a model with a factor of two crustal-thickness variation (Fig. 3b) is shown in Fig. 4b. In this model we have chosen a large crustal-thickness variation to enhance the development of the fringing high. This thick, low-density crust beneath the plateau results in a significantly larger (by a factor of about four) component of isostatic topography compared to the variable-thickness lithosphere model. However, the calculated isostatic topography is higher than the observed elevation of the plateau, indicating that this crustal-thickness variation exceeds that which may be present beneath Lakshmi Planum. The smaller-amplitude topographic high at the plateau edge, compared to that in the variable-thickness lithosphere model, can be increased by introducing a weaker lower crust due to strain softening³⁴, higher temperatures at depth, or different compositions than we have assumed. To explain plateau-fringing topographic highs by crustal-thickness changes alone would require thermal gradients significantly in excess of 20 K km^{-1} and a crustal rheology much weaker than expected³². We thus consider this mechanism to be an unlikely explanation of mountain belts in Ishtar Terra, though more detailed modelling will be necessary to assess its potential viability for fold-and-thrust belts in other areas.

Because of mesh resolution and the assumption of only large-scale ($10\text{--}10^2 \text{ km}$), smoothly varying lateral variations in lithospheric mechanical properties, this model does not address the nature of kilometre-scale short-wavelength fold-and-thrust belt deformation on Venus. Such deformation may be a consequence of imbricate thrusting^{3,35} or folding of shallow layers^{31,36,37} with a thickness much less than the depth to the brittle-ductile transition.

Mountain-belt-fringed plateaux are also found on Earth, with prominent examples including the Himalayas at the edge of the Tibetan plateau, and the Andes at the western margin of the Altiplano. Future work should assess the extent to which a similar pattern of long-wavelength topography can be explained by completely different mechanisms, such as underthrusting associated with subduction³⁸, and marginal plateau erosion and iso-

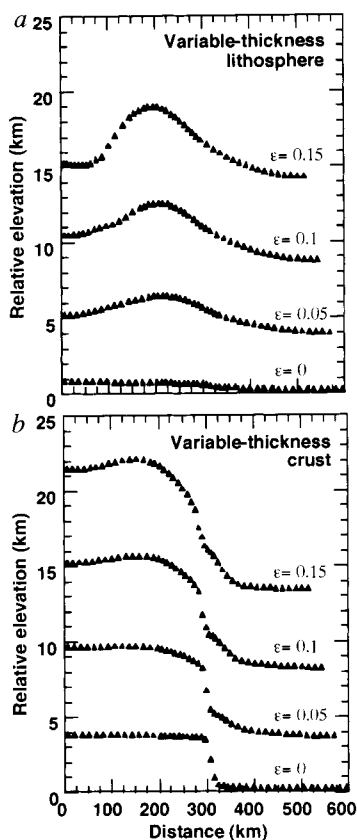


FIG. 4 Topographic profiles for variable-thickness lithosphere (a) and variable-thickness crust (b) models in Fig. 3 for strains $\epsilon = 0, 0.05, 0.1$ and 0.15 . Vertical exaggeration and horizontal scales are the same as in Fig. 1. Differences in the shapes of the calculated profiles and the idealized profile in Fig. 2 are due to the inclusion in the finite-element models of buoyancy forces and more complex lithosphere rheologies, as well as the assumption of finite strain.

static uplift³⁹, versus dynamic moments due to horizontal shortening of a laterally heterogeneous lithosphere. □

Received 3 May; accepted 15 September 1995.

- Campbell, D. B., Head, J. W., Harmon, J. K. & Hine, A. A. *Science* **221**, 647–664 (1983).
- Crumpler, L. S., Head, J. W. & Campbell, D. B. *Geology* **14**, 1031–1034 (1986).
- Head, J. W. *Geology* **18**, 99–102 (1990).
- Solomon, S. C. et al. *Science* **252**, 297–312 (1991).
- Solomon, S. C. et al. *J. geophys. Res.* **97**, 13199–13255 (1992).
- Vorder Bruegge, R. W. & Fletcher, R. C. *Lunar planet. Sci. Conf.* **21**, 1278–1279 (1990).
- Suppe, J. & Connors, C. J. *geophys. Res.* **97**, 13545–13561 (1992).
- Williams, C. A., Connors, C., Dahlen, F. A., Price, E. J. & Suppe, J. J. *geophys. Res.* **99**, 19947–19974 (1994).
- Price, R. A. in *Thrust and Nappe Tectonics* (eds McClay, K. R. & Price, N. J.) 427–448 (Geol. Soc., London, 1981).
- Cook, F. A. et al. *Geology* **7**, 563–567 (1979).
- Harris, L. & Bayer, K. C. *Geology* **7**, 568–572 (1979).
- Burchfiel, B. C. et al. *Geology* **17**, 448–452 (1989).
- Burchfiel, B. C. & Royden, L. H. *Eclologiae geol. Helv.* **84**, 599–629 (1991).
- Davis, D., Suppe, J. & Dahlen, F. A. *J. geophys. Res.* **88**, 1153–1172 (1983).
- Hoffman, P. F. in *Geology of North America—An Overview, Decade of North American Geology* (eds Baily, A. W. & Palmer, A. R.) 447–512 (Geol. Soc. Am., Denver, 1989).
- Munoz, J. A., Martinez, A. & Verges, J. J. *struct. Geol.* **8**, 399–406 (1986).
- Phillips, R. J. & Hansen, V. L. *Rev. Earth planet. Sci.* **22**, 597–654 (1994).
- Hansen, V. L. *Lunar Planet. Sci. Conf.* **23**, 478–479 (1992).
- Kuala, W. M. et al. *J. geophys. Res.* **97**, 16085–16120 (1992).
- Keep, M. & Hansen, V. L. *J. geophys. Res.* **99**, 26015–26028 (1994).
- Keep, M. & Hansen, V. L. *Lunar planet. Sci. Conf.* **25**, 681–682 (1994).

- Hansen, V. L. & Phillips, R. J. *Geology* **23**, 292–296 (1995).
- Grimm, R. E. & Phillips, R. J. *J. geophys. Res.* **96**, 8305–8324 (1991).
- Bathe, K. J. & Wilson, E. L. *Numerical Methods in Finite Element Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1976).
- Zuber, M. T. & Parmentier, E. M. *J. geophys. Res.* (in the press).
- Chen, Y. & Morgan, W. J. *J. geophys. Res.* **95**, 17583–17604 (1990).
- Desai, C. S. & Abel, J. F. *Introduction to the Finite Element Method* (Van Nostrand Reinhold, New York, 1972).
- Bindschadler, D. L., Schubert, G. & Kuala, W. M. *Geophys. Res. Lett.* **17**, 1345–1348 (1990).
- Lenardic, A., Kuala, W. M. & Bindschadler, D. L. *J. geophys. Res.* **98**, 18697–18705 (1993).
- Namiki, N. & Solomon, S. C. *J. geophys. Res.* **98**, 15025–15031 (1993).
- Zuber, M. T. *J. geophys. Res.* **92**, E541–E551 (1987).
- Mackwell, S. J., Zimmerman, M. E., Kohlstedt, D. L. & Scherber, D. S. in *Proc. 35th US Symp. on Rock Mechanics* (eds Daemen, J. J. K. & Schultz, R. A.) 207–214 (Balkema, Lake Tahoe, NV, 1995).
- Hansen, V. L. & Phillips, R. J. *Lunar planet. Sci. Conf.* **24**, 603–604 (1993).
- Zuber, M. T. *Lunar planet. Sci. Conf.* **25**, 1575–1578 (1994).
- Vorder Bruegge, R. W. & Head, J. W. *Geology* **19**, 885–888 (1991).
- Banerdt, W. B. & Golombek, M. P. *J. geophys. Res.* **93**, 4759–4772 (1988).
- Zuber, M. T. & Parmentier, E. M. *Icarus* **85**, 290–308 (1990).
- Kono, M., Fukao, Y. & Yamamoto, A. *J. geophys. Res.* **94**, 3891–3905 (1989).
- Masek, J., Isacks, B. L., Gubbels, T. L. & Fielding, E. J. *J. geophys. Res.* **99**, 13941–13956 (1994).
- Ford, P. G. & Pettengill, G. H. *J. geophys. Res.* **97**, 13103–13114 (1992).
- Parmentier, E. M. *Lunar planet. Sci. Conf.* **17**, 648–649 (1986).
- Grimm, R. E. *J. geophys. Res.* **99**, 23163–23171 (1994).

ACKNOWLEDGEMENTS. We thank R. Phillips and V. Hansen for detailed reviews, J. Garvin for making available his online database of Magellan topography, and S. Sakimoto for assistance with Figure 1. This work was supported by the NASA Planetary Geology and Geophysics Program and the NASA Solid Earth Sciences Program.

Assembly of submicrometre ferromagnets in gallium arsenide semiconductors

Jing Shi*, James M. Kikkawa*, Roger Proksch*, Tilman Schäffer*, David D. Awschalom*, Gilberto Medeiros-Ribeiro† & Pierre M. Petroff†

* Department of Physics, † Department of Materials and Electrical Engineering, University of California, Santa Barbara, California 93106, USA

THE discovery of spin-dependent electronic phenomena in magnetic multilayers¹ and granular solids^{2,3} has provided valuable insights into the nature of spin interactions in low-dimensional magnetic systems, and has opened the way to new technologies based on these phenomena. In the case of semiconductors, the incorporation of microscopic magnets would allow the electronic flexibility of semiconductor-based quantum structures to be combined with local magnetism⁴, potentially enabling the development of new tunable spin-dependent magneto-electronic and magneto-optical devices. Recent attempts to introduce ferromagnetism into III–V compound semiconductors have involved epitaxial growth of atomically thin layers, yielding two-dimensional magnetic films⁵ rather than localized magnetic structures. Here we describe a simple approach for fabricating discrete microscopic ferromagnets in the III–V semiconductor gallium arsenide. The semiconductor is first uniformly implanted with manganese ions. Subsequent heat treatment leads to a striking phase separation, whereby submicrometre crystals of GaMn nucleate and grow from the implanted layer. The resulting particles are ferromagnetic, with Curie temperatures exceeding room temperature.

The semiconductor samples are grown by molecular beam epitaxy on semi-insulating GaAs substrates, and consist of 500-nm undoped GaAs epitaxial layers grown atop 50-nm AlAs and a 500-nm GaAs buffer. These structures are then uniformly implanted with Mn⁺ ions at energies of 50 and 200 keV and doses ranging from 1×10^{14} to 5×10^{16} Mn⁺ cm⁻². To magnetically activate the implanted ions, the samples are covered with

Si substrates and annealed in a forming gas (90% N₂, 10% H₂) at temperatures from 800 to 920 °C for durations ranging from 5 s to 20 min. Both annealed and unannealed samples are thinned for transmission electron microscopy (TEM) studies. Electron microdiffraction and electron-beam-induced X-ray fluorescence spectroscopy are also employed for structural and compositional analysis. For magnetization measurements only, the GaAs substrate is removed to reduce the diamagnetic background. Etching the thin AlAs slab in diluted hydrofluoric acid releases the 500-nm-thick implanted surface from the substrate, and magnetization measurements of the resulting films are obtained from a radio frequency superconducting quantum interference device (SQUID) magnetometer in fields up to 5 T and temperatures ranging from 5 to 300 K. Atomic force microscopy (AFM) and room-temperature magnetic force microscopy (MFM) are employed to probe spatially both the topography of the sample surfaces and the local magnetic fields generated by the submicrometre magnetic structures.

The plan-view TEM image (Fig. 1a) reveals a striking phase separation in the Mn⁺-implanted and annealed GaAs samples which is absent in unannealed materials. The precipitates appear as dark spots and are typically 200 nm in diameter with an average separation of $\sim 1.5 \mu\text{m}$. The streaks in this image result from strain-induced bend-contours within the GaAs substrate. Semiquantitative analysis of the X-ray fluorescence spectra obtained from single precipitates in these structures indicates that they contain approximately 60% Ga and 40% Mn. Cross-sectional TEM micrographs, such as Fig. 1b (a sample implanted with 5×10^{16} Mn⁺ cm⁻²) reveal that most of the precipitates pierce the implanted surface. This may be due either to interfacing strain with the surrounding GaAs or point defects within the material, and allows monitoring of precipitate formation by AFM. Two-dimensional electron diffraction patterns were obtained from some of the GaMn particles. Analysis of the patterns indicates that some of these crystallites exhibit *m*- or fivefold symmetry in the icosahedral space group $\bar{m}\bar{3}5$ suggesting quasicrystalline order (J. P. Zhang, *et al.*, manuscript in preparation). As Ga is a neighbour of Al in the periodic table, similar quasicrystalline effects to those seen in AlMn systems⁶ may be expected.

To correlate precipitate formation with magnetic properties, we have studied both unannealed and annealed samples using

AFM and SQUID measurements. In the unannealed control sample receiving $5 \times 10^{16} \text{ Mn}^+ \text{ cm}^{-2}$, AFM data of Fig. 2a and cross-sectional TEM studies indicate no precipitate formation. As seen in Fig. 2c, the SQUID magnetometry at temperature $T = 5 \text{ K}$ shows that the implanted Mn^+ ions respond paramagnetically, as expected from single or low-spin complexes, and are well described by a Brillouin function. Subsequent annealing at 920°C for 60 s produces dramatic changes in both the structural and magnetic properties of the sample. AFM images reveal the appearance of precipitates (Fig. 2b), and the resulting magnetization contains a clear hysteresis loop superimposed on a superparamagnetic background (Fig. 2d). Comparison of the saturated ferromagnetic moments and the extrapolation of superparamagnetic moments to high fields ($\sim 20 \text{ T}$) using a Langevin fit indicates that 13% of the magnetically active Mn ions

are incorporated within the ferromagnetic particles. Moreover this ferromagnetism persists to room temperature, enabling sub-micrometre magnetic imaging by room-temperature MFM.

MFM measurements confirm that the ferromagnetism originates from the GaMn crystallites. For an annealed sample receiving $5 \times 10^{16} \text{ Mn}^+ \text{ cm}^{-2}$, comparison of the structural response seen in the topographic image of Fig. 3a and the simultaneous magnetic response appearing in Fig. 3b reveals a one-to-one correspondence between crystallites and the magnetic response of the MFM tip. Although most of the GaMn crystallites appear to be magnetic, they exhibit a variety of magnetic orientations and states. At this dosage, approximately 50% of the observed structures are clearly ferromagnetic; at a lower dosage of $1 \times 10^{15} \text{ Mn}^+ \text{ cm}^{-2}$ the ratio increases to 70%. Because the magnetic image contrast results from repulsive and active forces between the MFM tip and the crystallite, light and dark regions arise from oppositely oriented field gradients and correspond to opposite magnetic poles. The sample shown in Fig. 3 is in zero applied field and the precipitates generally appear to be dipolar with random orientations. A typical in-plane dipole is seen in Fig. 3 inset. In applied fields, switching of these individual particles is observed¹².

Direct SQUID magnetization measurements of these materials reveal that the GaMn crystallites are robust room-temperature ferromagnets. In Fig. 4a we see that strong magnetic hysteresis behaviour exists at 300 K. The coercive field, H_c ,

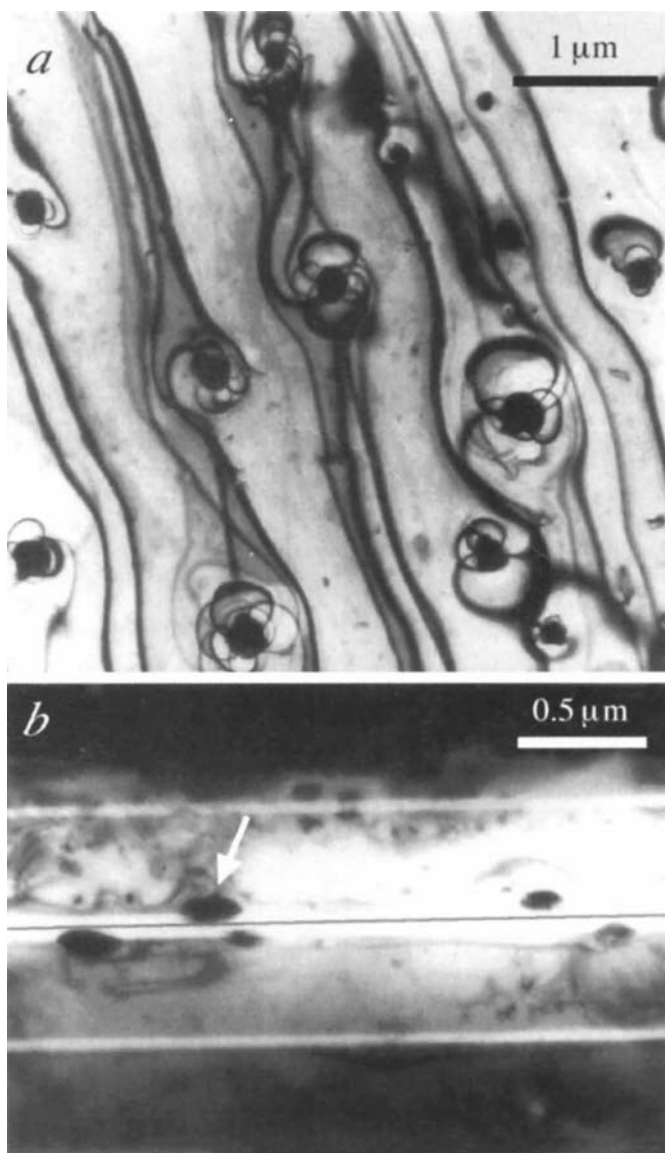


FIG. 1 *a*, Plan-view TEM micrograph of a sample implanted with $1 \times 10^{16} \text{ Mn}^+ \text{ cm}^{-2}$ at 200 keV and annealed at 920°C for 60 s. The dark areas are GaMn precipitates. *b*, Cross-sectional TEM micrograph of a split $5 \times 10^{16} \text{ Mn}^+ \text{ cm}^{-2}$ sample annealed at 920°C for 60 s. The sample in *b* was prepared by contacting the two implanted surfaces before thinning, and the two resulting cross-sections are shown with the boundary between surfaces marked by a solid line. An arrow points to one ferromagnetic precipitate.

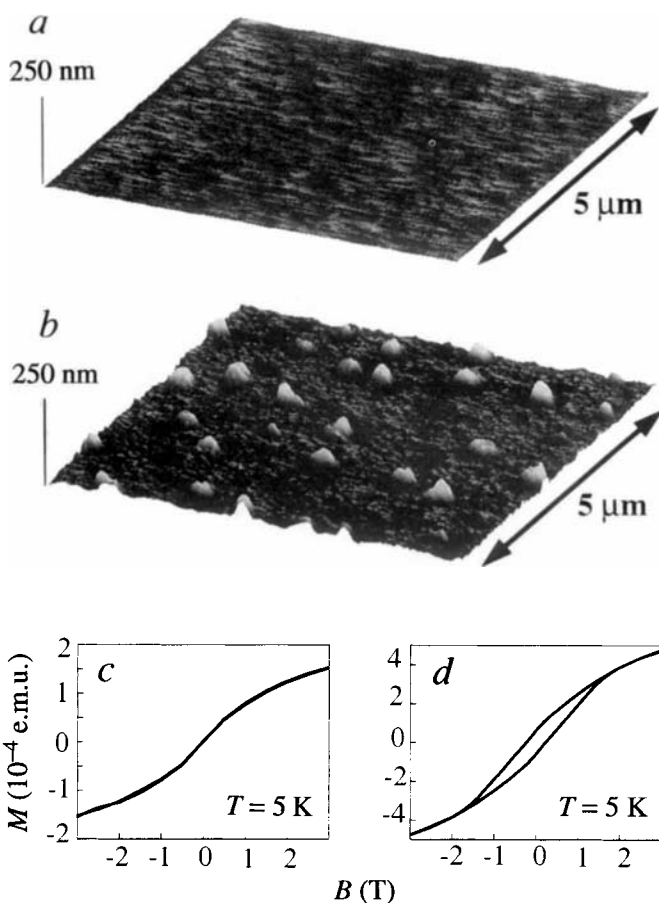


FIG. 2 AFM images of a sample implanted with $5 \times 10^{16} \text{ Mn}^+ \text{ cm}^{-2}$; *a*, unannealed and *b*, annealed at 920°C for 60 s. Paramagnetic magnetization for the unannealed sample is shown in *c*, and the superparamagnetic and ferromagnetic contributions for the annealed sample are shown in *d*. M is the magnetic moment, and B is the applied field in tesla.

(~ 0.5 T) changes little over the entire temperature range (5–300 K). In addition, the remanent magnetization (Fig. 4b) drops by only 15% from 5–295 K, implying a Curie temperature exceeding room temperature and at or above the Curie temperatures of other known GaMn (~ 450 –800 K)^{7,8} or MnAs (~ 320 K)⁹ compounds. For comparison, Fig. 4b demonstrates that most of the temperature-dependent magnetization arises from the superparamagnetic contribution.

Varying implantation dose and energy can affect precipitation by controlling the density of both Mn ions and lattice defects. Though a reduction of defects should lower Mn diffusion rates,

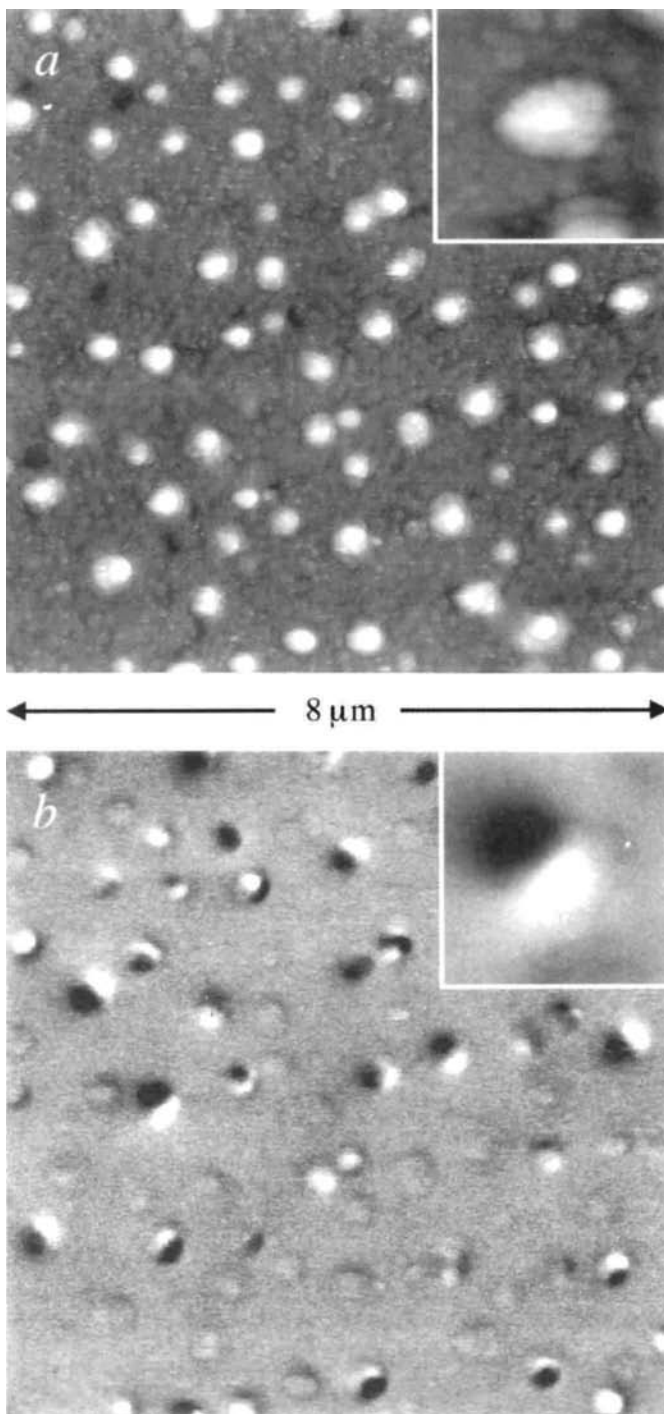


FIG. 3 Zero-field topographic (a) and magnetic-force (b) images of the same area taken on an unmagnetized $5 \times 10^{16} \text{ Mn}^+ \text{ cm}^{-2}$ sample annealed at 920°C for 60 s. The insets show enlarged structural and magnetic images of a single precipitate.

we find that reducing the dose by a factor of 50 (from 5×10^{16} to $1 \times 10^{15} \text{ Mn}^+ \text{ cm}^{-2}$) at a fixed annealing condition of 920°C for 60 s results in increased precipitation per ion. As imaged by AFM, the particle diameter, d , more than doubles from ~ 181 to ~ 400 nm, whereas precipitate densities only decrease to one-fifth their value ($1.5 \times 10^8 \text{ cm}^{-2}$ to $2.8 \times 10^7 \text{ cm}^{-2}$). Furthermore, the fractional size variation, $\Delta d/\bar{d}$, decreases from 0.91 to 0.71. (Here $\bar{d}$ is the r.m.s. value of d .) Magnetic measurements reflect these structural changes, where enhanced precipitation gives rise to nearly the same total magnetization with lower dosage. Moreover, H_c drops from 0.5 T to 0.1 T while the magnetization slope at H_c doubles, reflecting increased particle sizes and uniformities, respectively, and perhaps differences in particle profiles and compositions. We note that within the parameters studied, these magnetic properties were insensitive to the implantation energy and annealing conditions.

An improved understanding of implantation and annealing conditions as they relate to the size, shape, crystallinity and magnetism of these precipitates should allow tailoring of their individual magnetic properties and control of their spatial distributions. Furthermore, similar phase separations may occur in other semiconducting systems of current technological interest. For example, we observe both superparamagnetism and ferro-

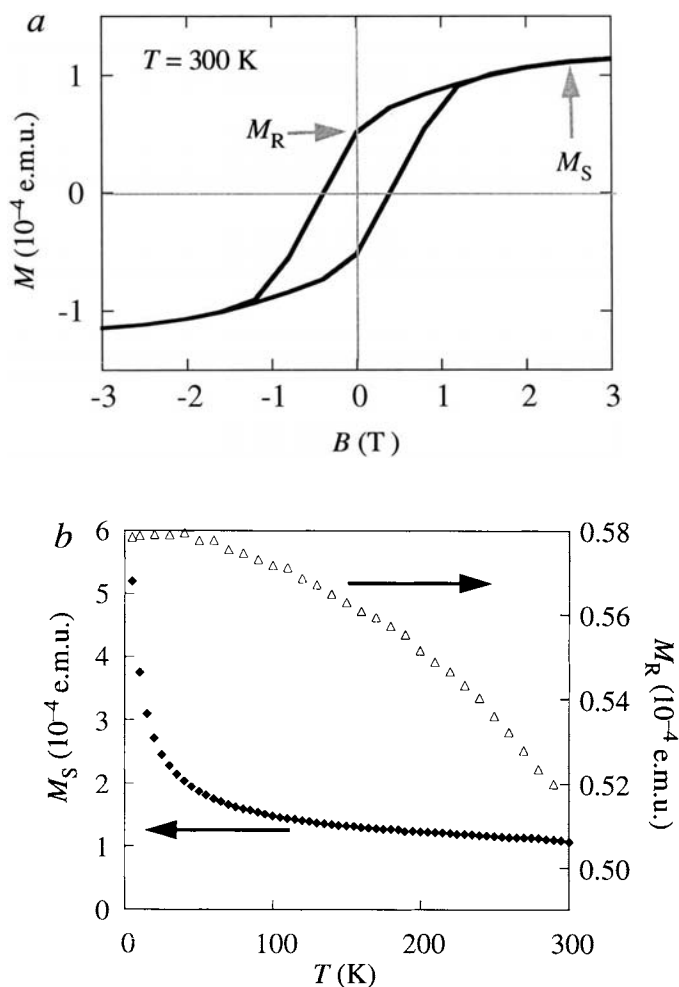


FIG. 4 a, SQUID data of a sample implanted with $5 \times 10^{16} \text{ Mn}^+ \text{ cm}^{-2}$ and annealed at 920°C for 60 s. Temperature is 300 K and the diamagnetic background is removed. M_R is the remanent moment at zero magnetic field, and M_S is the magnetization at $B = 2.5$ T, equal to the saturation moment of the ferromagnetic component at $T = 300$ K, where superparamagnetic contributions are small. b, Temperature dependence of M_R and M_S .

magnetism at temperatures up to 100 K in Si upon co-implantation with Mn^+ and Ga^+ . It is also likely that a focused ion beam, previously used to locally implant Mn ions¹⁰, could be employed to pattern ferromagnetic particles within existing semiconductor electronics. These techniques offer tantalizing possibilities for integrating a variety of submicrometre ferromagnetic structures¹¹ within III–V semiconductors, thus producing a new family of compact spin-dependent magnetoelectronic materials. □

Received 13 April; accepted 18 May 1995.

1. Parkin, S. S. P., Bhadra, R. & Roche, K. P. *Phys. Rev. Lett.* **66**, 2152–2155 (1991).
2. Berkowitz, A. E., Mitchell, J. R., Carey, M. J. & Young, A. P. *Phys. Rev. Lett.* **68**, 3745–3748 (1992).
3. Xiao, J. Q., Jiang, J. S. & Chien, C. L. *Phys. Rev. Lett.* **68**, 3749–3752 (1992).
4. Prinz, G. *Science* **250**, 1092–1097 (1990).
5. Tanaka, M. *et al. Appl. Phys. Lett.* **63**, 696–698 (1993).
6. Shechtman, D., Blech, I., Gratias, D. & Cahn, J. W. *Phys. Rev. Lett.* **53**, 1951–1953 (1984).
7. Tsuboya, I. & Sugihara, M. *J. phys. Soc. Japan* **18**, 1096 (1963).
8. Wachtel, E. & Neiri, K. J. Z. *Metallkd.* **56**, 779–789 (1965).
9. Massalski, T. B., Okamoto, H., Subramanian, P. R. & Kacprzak, L. (eds) *Binary Alloy Phase Diagram* 2nd edn Vol. 1 295 (Am. Soc. Metals, Metals Park, OH, 1990).
10. Kikkawa, J. M., Baumberg, J. J., Awschalom, D. D., Leonard, D. & Petroff, P. M. *Phys. Rev. Rapid Commun.* **B50**, 2003–2006 (1994).
11. Pekarek, T. M. *et al. Appl. Phys. Lett.* (submitted).
12. J. Shi *et al. J. appl. Phys.* (submitted).

ACKNOWLEDGEMENTS. We thank J. P. Zhang for the electron diffraction measurements. This work was supported in part by the US Air Force Office of Scientific Research, the US NSF and the QUEST NSF Center for Quantized Electronic Structures.

Effects of clouds and stratospheric ozone depletion on ultraviolet radiation trends

Dan Lubin* & Elsa H. Jensen†

* California Space Institute, University of California, San Diego, La Jolla, California 92093-0221, USA

† SeaSpace Corporation, San Diego, California 92126, USA

ANTHROPOGENIC depletion of ozone in the lower stratosphere has been of global environmental concern for two decades, but the environmentally relevant quantity—the flux of solar ultraviolet radiation (UVR) reaching the Earth's surface—remains poorly quantified on a global basis. The three most important parameters governing surface UVR fluxes and trends are solar elevation, total vertically integrated ozone abundance and cloud opacity. Here we use global satellite measurements of total ozone abundance and cloud reflectance to examine how the trends in UVR resulting from established trends in total ozone abundance^{1,2} compare with the potentially large natural variability in UVR that results from variations in cloud opacity. We find that throughout many temperate regions—including large parts of continental Europe, North and South America, New Zealand, Australia and southern Africa—interannual variability in cloud opacity is sufficiently small that by the end of this century, trends in summer average local-noon UVR dose rates relevant to mammalian skin cancer or plant damage should be significant with respect to cloud variability.

We mapped the fluxes of local-noon UVR over most inhabited areas by using data from NASA's Total Ozone Mapping Spectrometer (TOMS, version 6 data)³ and the Earth Radiation Budget Experiment satellites (ERBE, S-9 monthly/hourly scanner data)⁴ as input to radiative transfer models based on the delta-Eddington approximation⁵. After co-location of the TOMS and ERBE data, the three-step mapping procedure yielded a spatial resolution (pixel size) of $2.5^\circ \times 2.5^\circ$ (latitude $\times$ longitude). First, the ERBE S-9 monthly/hourly clear-sky short-wave (wavelength integral 0.28–4.0 μm) albedos were used to estimate the surface albedo following the method of Li and

Garand⁶. Second, the shortwave radiation algorithm from the National Center for Atmospheric Research Community Climate Model (CCM2)⁷ was used to estimate an effective cloud optical depth from the ERBE S-9 monthly/hourly shortwave planetary albedos, along with the above surface albedo retrievals and the monthly averaged total column abundances from TOMS. For both steps, total column water vapour abundance was measured by the Tiros Operational Vertical Sounder (TOVS) instrument aboard the NOAA polar orbiting satellites⁸. Third, with the above retrievals and ozone abundance from TOMS, a spectrally resolved delta-Eddington algorithm was used to estimate monthly average local-noon UVR flux at the Earth's surface⁹. These spectral fluxes were then weighted by several commonly used action spectra to yield dose rates¹⁰; the dose rates we report here are those pertaining to mammalian skin cancer¹¹ and damage to plant DNA¹². This procedure for estimating surface UVR from satellite data is not the only one possible^{9,13,14}, but has the advantage that each of the satellite measurements (ozone, cloud reflectance and water vapour) was made by an orbiting instrument specifically designed to make the respective measurement.

The five-year local-noon UVR fluxes derived as above exhibit a great deal of spatial and temporal variability due to cloud cover (Fig. 1), but also implicitly include the recently established trends in ozone². Here we wish to separate and contrast the effects of ozone and cloud cover. To accomplish this, we repeated the delta-Eddington calculations of surface UVR flux for each month and year of the ERBE cloud effective optical depth retrievals, but using only the TOMS ozone from 1985 for each respective month. This allowed us to estimate the variability in local-noon UVR that results only from interannual variability in cloud opacity. For each month, a sample standard deviation in the dose rate, σ_c , was computed for each pixel from this second set of delta-Eddington calculations. These standard deviations show considerable zonal and regional variability (Fig. 2).

We wish to determine (for a given month and location) the number of years for which we must steadily deplete the ozone layer before the resulting increase in UVR local-noon dose rate becomes significant above the background variability due to interannual variability in cloud cover. To address this, we computed a baseline delta-Eddington UVR climatology using the 1985 TOMS ozone and the mean cloud effective optical depth from the five years of ERBE retrievals. Then in a computer simulation for each month and pixel, we allowed the 1985 TOMS ozone to be depleted (or in some cases increased) everywhere according to the regional and monthly trends reported by Niu *et al.*². As the ozone was depleted in one-year increments, the surface local-noon UVR dose rate, F_d , was recomputed for each pixel. At each pixel, we noted the number of years that elapsed before the dose rate increased by a threshold of $1.96\sigma_c$. This elapsed time is mapped for the skin-cancer dose rate in Fig. 3 and the plant-damage dose rate in Fig. 4. If we take the onset of ozone depletion to be *circa* 1970, then all regions shown in orange or yellow should already be experiencing (or should shortly experience) upward trends in the local-noon dose rate that are significant relative to interannual variability in cloud opacity.

In a Northern Hemisphere summer month (July) and for the skin-cancer dose rate, most of continental Europe and the Mediterranean should be showing UVR increases above the threshold of cloud variability within 10–30 years after the onset of ozone depletion. But, for much of the UK and Ireland, increases in the skin-cancer dose rate would not be significant against cloud variability until 30–60 years after the onset of ozone depletion. Other regions where increases in the skin-cancer dose rate would not be significant against cloud variability for more than 40–100 years after the onset of ozone depletion include much of the Caucasian steppe, a region in central Russia east of Moscow, most of China and the Indian subcontinent, and most of Japan

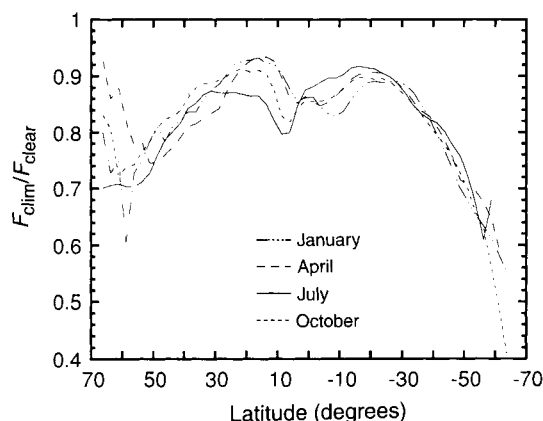


FIG. 1 Examples of the effect of cloud opacity on the total UV-B (280–315 nm) flux at the Earth's surface, for the year 1989. Shown here is the ratio of the zonally averaged surface fluxes from our climatology (F_{clim}) to the zonally averaged fluxes that would prevail under clear skies, F_{clear} (all other atmospheric parameters held constant). The effective cloud optical depth¹⁸ used to map the surface UVR fluxes is defined as $\tau = 1.5\text{LWP}/r_e$ where LWP is the cloud liquid water path and r_e is the effective radius of the cloud particle-size distribution²². At each pixel, τ is estimated from the CCM2 radiation algorithm by iterating over LWP until the modelled top-of-atmosphere solar flux matches the ERBE measurement. Values of r_e are fixed at 7 μm over land and 11 μm over ocean. The CCM2 radiation algorithm was modified to account for the 'excess cloud absorption' phenomenon²³ by using a parametrization provided by J. Kiehl (personal communication). For a given ERBE flux measurement, this modification has the effect of increasing the estimated τ and decreasing the mapped UVR flux. For most of the ERBE data analysed, this decrease in the UVR flux (relative to the case of no 'excess cloud absorption') is <20%, but the change becomes larger (up to 50%) for large ERBE flux measurements under high Sun elevation. The existence of the 'excess cloud absorption' phenomenon (still controversial²⁴) therefore has some bearing on this study. The overall behaviour of the flux ratios in this figure is, however, consistent with previous satellite studies based on different data¹³, and the behaviour of the flux ratios during October at southern high latitudes is consistent with field UV-B measurements¹⁷.

and Korea. Throughout the United States during July, regions where increases in the skin-cancer dose rate become significant against cloud variability within 30 years include the mid-west, southwest and Hawaii; regions where this would not occur before 30–50 years have elapsed include the southern gulf states, the Pacific northwest, and much of the Alaskan coastline. For most of Canada, with the exception of Pacific coastal areas and part of central Ontario, increases in the skin-cancer dose rate should exceed the threshold of interannual cloud variability within 30 years after the onset of ozone depletion. This is also true for the Caribbean, although longer elapsed times are indicated for much of Mexico and Central America. During a Southern

Hemisphere summer month (January), regions where increases in the skin-cancer dose rate would exceed the threshold of cloud-induced interannual variability within 30 years include the southern half (and most populated section) of Australia, all of New Zealand, most of South Africa, and the southern third of South America including Santiago and Buenos Aires, but not as far north as Sao Paulo. The regional behaviour in the skin-cancer dose rate is mirrored in the plant-damage dose rate, with the important exception that the elapsed times are generally a factor of two longer for the plant-damage dose rate due to the larger UV-A weighting in the plant-damage action spectrum¹².

FIG. 2 Standard deviations σ_c in the unweighted local-noon UV-B flux that result from interannual variability in cloud opacity as estimated from the ERBE data, for July. Standard deviations for all dose rates are very similar to those for the unweighted UV-B surface flux. Although the ERBE data are a small sample from which to estimate interannual variability, its five-year duration did encompass one pronounced El Niño–Southern Oscillation event²⁵ and these data should therefore provide a useful first-order estimate of global variability on decadal timescales. These standard deviations, as well as the trends in ozone derived by Niu *et al.*², show enough geographical variability from one month to the next that we present results for a monthly average rather than a seasonal average.

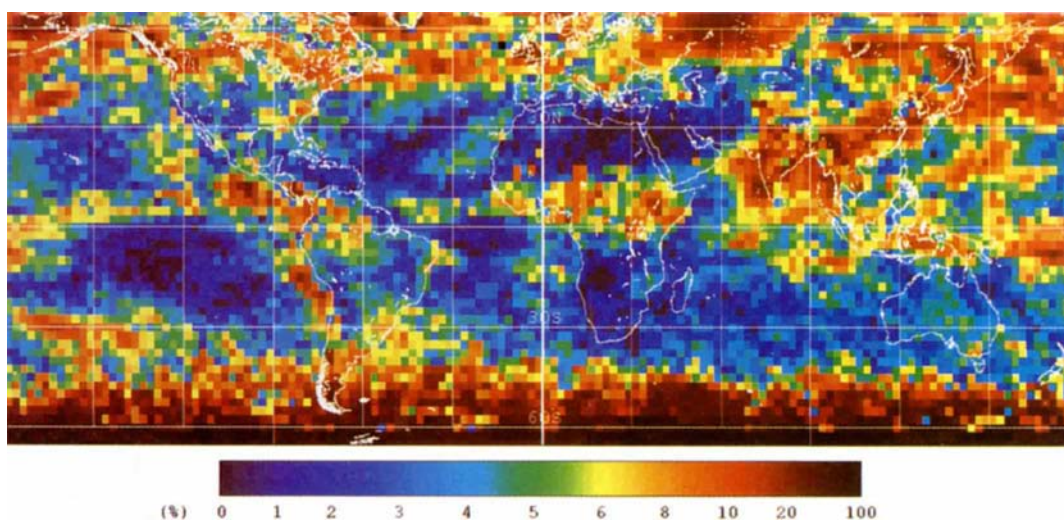
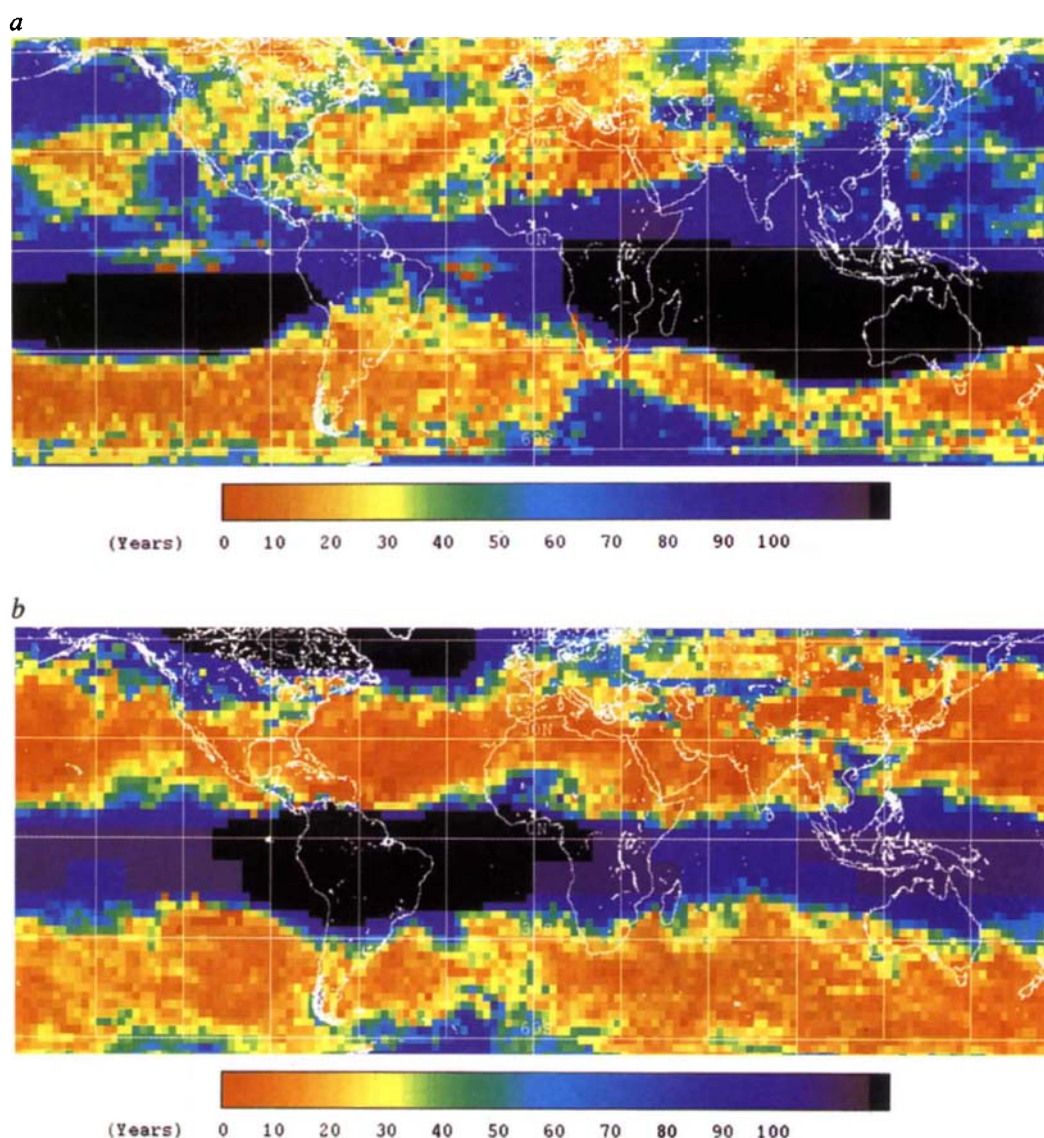
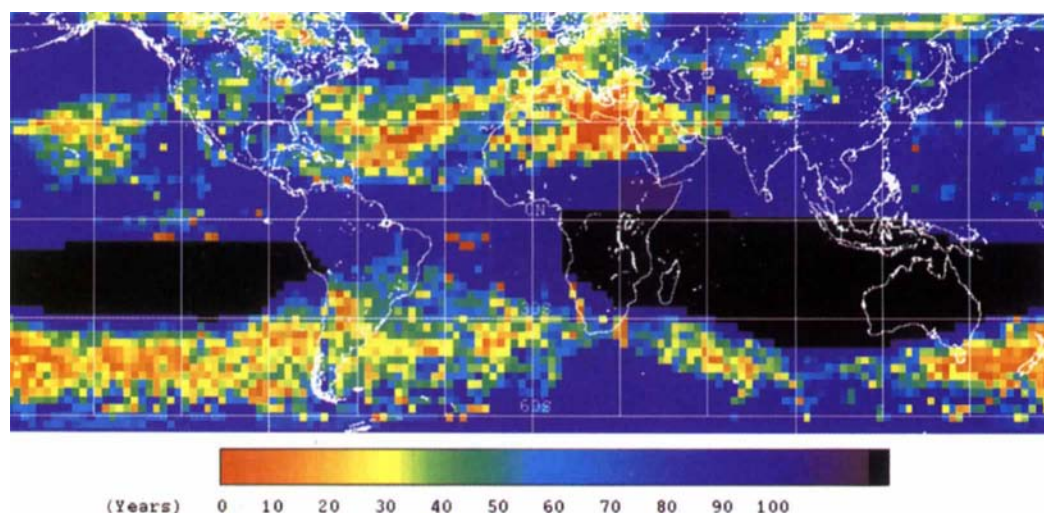


FIG. 3 Global map of the elapsed time in years after the onset of ozone depletion before the climatological local-noon dose rate for skin cancer has increased by 1.96 standard deviations of the cloud-induced interannual variability; *a*, for July, and *b*, for January. In choosing this threshold, we assume that the interannual variability in cloud opacity can be represented by a normal distribution, and that it is approximately constant over the timescale of the ozone-depletion issue. We assume that at time t_0 before ozone depletion, the climatological UVR dose rate in a given month and location can be characterized by a mean, $\langle F_{d0} \rangle$ (defined by both mean cloud opacity and total ozone), and by a standard deviation σ_c due to interannual variability in cloud opacity. After the onset of ozone depletion, the mean dose rate $\langle F_d \rangle$ will change according to a secular trend that we can estimate with our radiative-transfer model given an estimate of the trend in ozone. If at some time $t > t_0$ we make an estimate of the dose rate F_d for a given pixel, then we can say with a 95% confidence level that the true climatological mean $\langle F_d \rangle$ lies within $\pm 1.96\sigma_c$ of that estimate. If, at time t , the interval $F_d \pm 1.96\sigma_c$ no longer encompasses the initial mean $\langle F_{d0} \rangle$, then we say that at time t the trend in $\langle F_d \rangle$ has become significant above the interannual variability in cloud opacity, and that the UVR increase might therefore be a biological stress. This is not a statistical definition for UVR trend detection; it is instead a (more strict) criterion



for biological significance of a UVR trend. The darkest blue and violet areas refer to regions where the UVR change is either zero or slightly negative.

FIG. 4 As in Fig. 3, but for the local-noon dose rate for plant DNA damage and the month of July. We caution that there is a wide variability in reported action spectra for plant damage^{10,12}, and that these results pertain only to one of the widely used plant action spectra; we also note that Figs. 2–4 should be interpreted on a regional (mesoscale) basis only.



The results presented here are a first attempt at investigating expected global trends in surface UVR in the context of cloud variability, and we now give several cautionary notes. First, the cloud reflectance dataset (ERBE) has a duration of only five years, and could therefore provide a first-order estimate of inter-annual variability, but not any estimate of possible trends in the global cloud cover itself. Second, the global trends in ozone used here² were derived from TOMS data over the period 1978–91. There have been some pronounced ozone depletion events during the early 1990s¹⁵, but on the other hand, ozone depletions are expected to recover during the first half of the next century as a result of the 1987 Montreal Protocol. Certainly this type of study should be repeated periodically, as the TOMS or similar ozone time series becomes longer and the regional ozone trends are revised accordingly. Third, in this study tropospheric ozone is held constant. Tropospheric ozone can provide a noticeable absorption of UV-B (280–315 nm) flux in the presence of cloud scattering¹⁶, and for regions where tropospheric ozone abundance may be increasing (for example, industrialized regions), the elapsed times reported here might best be considered as lower limits. Fourth, satellite investigations of this kind should not be taken as a substitute for deploying a global ground-based UVR monitoring network using spectral radiometric equipment. Fifth, we have examined the significance of UVR trends with respect to cloud variability because the issue of cloud opacity has hitherto been considered an unknown quantity on a global basis. We have not discussed detection of UVR trends with respect to variability in ozone; this is beyond the scope of this study and is best addressed by ground-based spectral UVR measurements^{17–19}. There are some reasons for being concerned about peak (cloud-free) rather than climatological UVR exposure^{20,21} (for example, the effects of recreational sun exposure in humans), and in the cloud-free case our threshold for biological significance is reached immediately. Last, we have discussed here only the elapsed time from the onset of ozone depletion to when a trend in a given biological UVR dose rate would just begin to exceed the bounds under which ecosystems have adapted, as defined by interannual variability in cloud opacity. The ecological significance of this should be determined by the workers in the area of photobiology. □

The internal structure of an active sea-floor massive sulphide deposit

Susan E. Humphris¹, P. M. Herzig², D. J. Miller³, J. C. Alt⁴, K. Becker⁵, D. Brown⁶, G. Brügmann⁷, H. Chiba⁸, Y. Fouquet⁹, J. B. Gemmell¹⁰, G. Guerin¹¹, M. D. Hannington¹², N. G. Holm¹³, J. J. Honnorez¹⁴, G. J. Iturrino⁵, R. Knott¹⁵, R. Ludwig¹⁶, K. Nakamura¹⁷, S. Petersen², A.-L. Reysenbach¹⁸, P. A. Rona¹⁹, S. Smith²⁰, A. A. Sturz²¹, M. K. Tivey²² & X. Zhao²³

THE hydrothermal circulation of sea water through permeable ocean crust results in rock–water interactions that lead to the formation of massive sulphide deposits. These are the modern analogues of many ancient ophiolite-hosted deposits^{1–4}, such as those exposed in Cyprus. Here we report results obtained from drilling a series of holes into an actively forming sulphide deposit on the Mid-Atlantic Ridge. A complex assemblage of sulphide–anhydrite–silica breccias provides striking evidence that such hydrothermal mounds do not grow simply by the accumulation of sulphides on the sea floor. Indeed, the deposit grows largely as an *in situ* breccia pile, as successive episodes of hydrothermal activity each form new hydrothermal precipitates and cement earlier deposits. During inactive periods, the collapse of sulphide chimneys, dissolution of anhydrite, and disruption by faulting cause brecciation of the deposit. The abundance of anhydrite beneath the present region of focused hydrothermal venting reflects the high temperatures (>150 °C) currently maintained within the mound, and implies substantial entrainment of cold sea water into the interior of the deposit. These observations demonstrate the important role of anhydrite in the growth of massive sulphide deposits, despite its absence in those preserved on land.

In September–November 1994, the Ocean Drilling Program Leg 158 drilled a series of holes into a large hydrothermally active mound in the TAG hydrothermal field located at a water depth of 3,650 m on the eastern side of the median valley of the Mid-Atlantic Ridge at 26° 08' N (Fig. 1 inset). Although hydrothermal deposits have been found over an area of about 5 × 5 km, known high-temperature venting is at present confined to the TAG active mound which is about 200 m in diameter and 50 m high^{5–7}. Episodic hydrothermal activity at this site over at least the past 20,000 years^{8,9} has resulted in the construction of two superposed platforms¹⁰. Mineralogical zoning on the surface of the mound reflects different, but related, types of venting (Fig. 1)^{11,12}. Chalcopyrite–anhydrite-rich chimneys venting high-

Received 8 May; accepted 21 September 1995.

- Bojkov, R., Bishop, L., Hill, W. J., Reinsel, G. C. & Tiao, G. C. *J. geophys. Res.* **95**, 9785–9807 (1990).
- Niu, X., Frederick, J. E., Stein, M. L. & Tiao, G. C. *J. geophys. Res.* **97**, 14661–14669 (1992).
- Stolarski, R. S., Bloomfield, P. & McPeters, R. D. *Geophys. Res. Lett.* **18**, 1015–1018 (1991).
- Ramanathan, V. et al. *Science* **243**, 57–63 (1989).
- Joseph, J. H., Wiscombe, W. J. & Weinman, J. A. *J. atmos. Sci.* **33**, 2453–2459 (1976).
- Li, Z. & Garland, L. *J. geophys. Res.* **99**, 8335–8350 (1994).
- Briegleb, B. P. *J. geophys. Res.* **97**, 7603–7612 (1992).
- Wittmeyer, I. L. & Vonder Haar, T. H. *J. Clim.* **7**, 326–333 (1994).
- Lubin, D., Ricchiazzi, P., Gautier, C. & Whritner, R. H. in *Ultraviolet Radiation in Antarctica: Measurements and Biological Effects* (eds Weiler, S. & Penhale, P.) 53–81 (Antarctic Res. Ser. Vol. 62, Am. Geophys. Union, Washington DC, 1994).
- Madronich, S. in *Environmental UV Photobiology* (eds Young, A. R., Björn, L. O., Moan, J. & Nultsch, W.) 1–39 (Pierum, New York, 1993).
- Cole, C. A., Forbes, D. & Davies, R. E. *Photochem. Photobiol.* **43**, 275–284 (1986).
- Quaife, F. E., Sutherland, B. M. & Sutherland, J. C. *Nature* **358**, 576–578 (1992).
- Frederick, J. E. & Lubin, D. *J. geophys. Res.* **93**, 3825–3832 (1988).
- Eck, T. F., Bhartia, P. K. & Kerr, J. B. *Geophys. Res. Lett.* **22**, 611–614 (1995).
- Gleason, J. F. et al. *Science* **260**, 523–526 (1993).
- Bruhn, C. & Crutzen, P. J. *Geophys. Res. Lett.* **16**, 703–706 (1989).
- Lubin, D. et al. *J. geophys. Res.* **97**, 7817–7828 (1992).
- Stamnes, K., Slusser, J., Bowen, M., Booth, C. R. & Lucas, T. *Geophys. Res. Lett.* **17**, 2181–2184 (1990).
- Seckmeyer, G. et al. *Geophys. Res. Lett.* **22**, 1889–1892 (1995).
- Madronich, S. *Geophys. Res. Lett.* **19**, 37–40 (1992).
- Blumthaler, M. & Ambach, W. *Science* **248**, 206–208 (1990).
- Liou, K. N. *Radiation and Cloud Processes in the Atmosphere* 305–307 (Oxford Univ. Press, 1992).
- Pilewskie, P. & Valero, F. P. *J. Science* **267**, 1626–1629 (1995).
- Li, Z., Barker, H. W. & Moreau, L. *Nature* **376**, 486–490 (1995).
- Ramanathan, V. & Collins, W. *Nature* **351**, 27–32 (1991).

ACKNOWLEDGEMENTS. We thank W. Collins and W. Hunter of the Center for Clouds, Chemistry and Climate for facilitating access to the ERBE data. This work was supported by NASA.

¹Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA; ²Institut für Mineralogie, Tu Bergakademie Freiberg, 09595 Freiberg, Germany; ³Ocean Drilling Program, Texas A&M University, College Station, Texas 77845, USA; ⁴Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA; ⁵Department of Marine Geology and Geophysics, Rosenstiel School of Marine and Atmospheric Sciences, University of Miami, Miami, Florida 33149, USA; ⁶Instituto de Ciencias de la Tierra, CSIC, Marti i Franques s/n 08028, Barcelona, Spain; ⁷Max-Planck Institut für Chemie, Abteilung Geochemie, Postfach 3060, D-55020 Mainz, Germany; ⁸Department of Earth and Planetary Sciences, Kyushu University 33, 6-10-1 Hakozaki, Fukuoka 812, Japan; ⁹IFREMER Centre de Brest, DRO/GM, BP 70 29280 Plouzané, France; ¹⁰CODES, University of Tasmania, GPO Box 252C, Hobart, Tasmania 7001, Australia; ¹¹Lamont-Doherty Earth Observatory, Palisades, New York 10964, USA; ¹²Geological Survey of Canada, 601 Booth Street, Ottawa, Ontario K1A 0E8, Canada; ¹³Department of Geology and Geochemistry, Stockholm University, S-10691, Stockholm, Sweden; ¹⁴Institut de Géologie, Université Louis Pasteur, 1 rue Blessig, 67084 Strasbourg Cedex, France; ¹⁵Department of Earth Sciences, University of Wales, Cardiff CF1 3YE, UK; ¹⁶SOEST, University of Hawaii at Manoa, 2525 Correa Road, Honolulu, Hawaii 96822, USA; ¹⁷Geological Survey of Japan, 1-1-3 Higashi, Tsukuba, Ibaraki 305, Japan; ¹⁸Department of Biology, Indiana University, Bloomington, Indiana 47405, USA; ¹⁹Institute of Marine and Coastal Sciences, Rutgers University, PO Box 231, New Brunswick, New Jersey 08903, USA; ²⁰Department of Geosciences, University of Houston, 4800 Calhoun, Houston, Texas 77204, USA; ²¹Marine and Environmental Studies Program, University of San Diego, Camino Hall 8C, Alcalá Park, San Diego, California 92110, USA; ²²Department of Marine Chemistry & Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA; ²³Institute of Tectonics, Department of Earth Sciences, University of California, Santa Cruz, California 95064, USA.

temperature ($>360^{\circ}\text{C}$) copper-rich fluids are clustered on a 20 m diameter cone on the northwestern side of the upper platform (the Black Smoker complex). Approximately 70 m to the southeast on the lower platform, a field (the Kremlin area) of sphalerite-dominated, white smokers vent zinc-rich fluids at temperatures of $260\text{--}300^{\circ}\text{C}$ (refs 13–15). Geochemical modelling of fluid compositions and the associated mineral assemblages indicates that the white-smoker fluids can be derived from the black-smoker fluids by a combination of conductive cooling and mixing with sea water, and precipitation of sulphides and anhydrite within the mound^{12,15}. These reactions decrease the pH of the circulating hydrothermal fluids to values below 3, causing dissolution of sphalerite within the mound and an order of magnitude enrichment of zinc in the white-smoker fluids (Zn concentration, $300\text{--}400\mu\text{mol l}^{-1}$) compared with the black-smoker fluids (Zn concentration, $46\mu\text{mol l}^{-1}$)¹⁵. The zinc is then reprecipitated at the surface of the mound as sphalerite-rich chimneys and domes. Diffuse flow of low-temperature ($<25^{\circ}\text{C}$) fluids occurs over the rest of the top and sides of the mound.

Seventeen holes drilled at five locations (TAG 1–5) on the TAG active mound reveal the subsurface nature and lateral heterogeneity of the actively forming deposit and the underlying upflow zone down to the depth of 125 metres below sea floor (m.b.s.f.) (Fig. 1). Because holes were drilled in close proximity (within 10–15 m) at each location, their stratigraphy has been combined to produce a composite section for each area (Fig. 2). Recovery of relatively unaltered basalt near the edges of the mound at depths of about 20 m.b.s.f. (at TAG-2) and 40 m.b.s.f. (at TAG-4) has constrained the lateral extent of intense crustal alteration and mineralized upflow zone to about 80 m. Based on the thickness of the sulphide zone and the extent of the underlying upflow zone, our preliminary estimates indicate that there are about 2.7 million tonnes of massive sulphides above the sea floor, and approximately 1.2 million tonnes below the sea floor in the upflow zone. This is well within the range for typical ophiolite-based massive sulphide deposits worldwide¹⁶ and is quite comparable with the average-sized Cyprus-type ore deposit of 3 million tonnes¹⁷. Shipboard analyses of samples from the drill core indicate bulk copper contents of 1–2 wt% for the deposit and imply a bulk metal content of 30–60 thousand tonnes of copper.

Breccias of various types dominate all the sections drilled through the mound (Fig. 2). Massive sulphides reflecting the formation of new hydrothermal precipitates (for example, massive granular pyrite and chalcopyrite at TAG-1, and massive sphalerite and pyrite at TAG-2) are restricted to the upper few metres at each drilling location. The near-surface precipitates at TAG-2 are relatively enriched in base metals (Zn contents of 1–4 wt%, Au contents of 3 p.p.m. compared with <700 p.p.m. and ~ 250 p.p.b. respectively at TAG-1); this is interpreted to be the result of remobilization of these metals at depth within the mound from previously deposited sulphides. Red and grey siliceous material (Fig. 3a) was also encountered in the upper few metres of the TAG-2, TAG-3 and TAG-4 areas. This probably results from the precipitation of silica at low temperatures ($<100^{\circ}\text{C}$) from hydrothermal fluids percolating through the mound.

Based on the overall internal structure of the mound and the upflow zone, four major zones can be distinguished, all of which may or may not be present in one section (Fig. 2). Zone 1 consists of clast-supported massive pyrite breccias that dominate the upper 10–20 m at every location (Fig. 3b). This is underlain by an anhydrite-rich zone (zone 2) identified only in the TAG-1 and TAG-5 areas. This zone can be divided into two breccia types distinguished primarily on the basis of the relative abundances of pyrite, anhydrite and silica: (1) matrix-supported pyrite-anhydrite breccias (Fig. 3c) are present down to about 30 m.b.s.f., and are underlain by (2) pyrite-silica-anhydrite breccias that extend to the bottom of the anhydrite-rich zone at about 45 m.b.s.f. Anhydrite veining is extremely well developed throughout this zone, where composite veins up to 45 cm thick are present (Fig. 3d). They comprise complex multi-stage fracture fillings and cavity linings, some of which include disseminated, fine-grained pyrite and chalcopyrite, and trace amounts of haematite. Zone 3 consists of intensely silicified and brecciated wall rock which comprises the upper portion of the upflow zone beneath the mound. Pyrite-silica breccias (Fig. 3e) occur in the upper part of this zone and are underlain by silicified wall-rock breccias, which are distinguished from the overlying pyrite-silica breccias in that they contain significantly less pyrite (<50 vol.%) and are dominantly clast-supported. Below about 100 m.b.s.f., the silicified wall-rock breccias grade into a chloritized basalt breccia (zone 4) where chloritized and weakly mineralized basalt

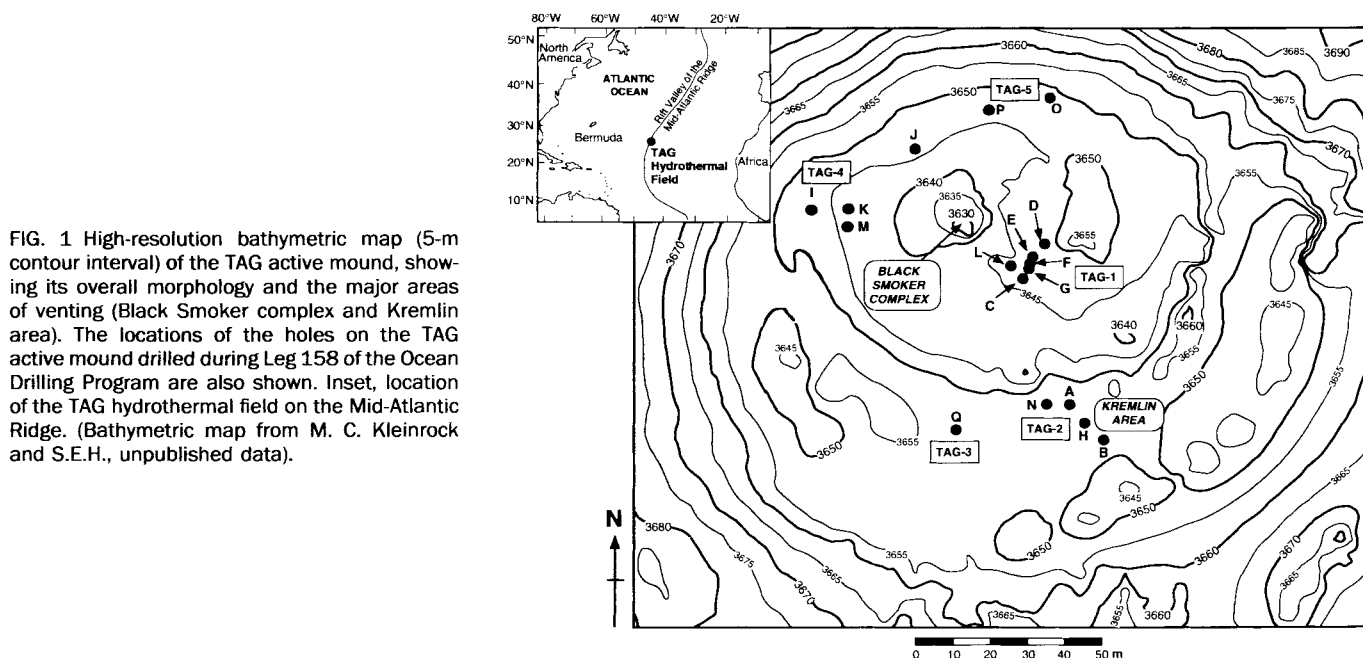


FIG. 1 High-resolution bathymetric map (5-m contour interval) of the TAG active mound, showing its overall morphology and the major areas of venting (Black Smoker complex and Kremlin area). The locations of the holes on the TAG active mound drilled during Leg 158 of the Ocean Drilling Program are also shown. Inset, location of the TAG hydrothermal field on the Mid-Atlantic Ridge. (Bathymetric map from M. C. Kleinrock and S.E.H., unpublished data).

fragments (1–5 cm in size) are cemented by quartz and pyrite, and are cross-cut by veins of pyrite, quartz and quartz + pyrite (Fig. 3f).

One of the most striking features of the cores recovered from the TAG mound is the dominance of breccias throughout the mound and in the upper parts of the underlying upflow zone. These breccias often include a mixture of clasts of various types (for example, sulphide, altered (silicified) wall rock) that formed by different processes at different depths. Of particular note is the presence of altered basalt fragments at shallow levels within the hydrothermal mound. This complex structure implies that the TAG active mound has undergone multiple stages of development, which are reflected in the sequences of alteration and veining events that can be distinguished in both the sulphide breccias and in the silicified and chloritized basalt breccias. The sulphide breccias probably accumulated at the sea floor by the collapse of large sulphide chimneys, mass wasting, and by dissection of massive sulphides by repeated movement along faults. This debris has been overgrown by later generations of deposits and progressively cemented or replaced by quartz, sulphides and sulphates. Closing and reopening of faults due to tectonic events could explain the focusing of fluid flow at this site during episodic hydrothermal activity. The presence of altered basalt clasts in the upper part of the mound may indicate (1) very high fluid flow rates¹⁸ that transport the clasts from depth and then re-deposit them at shallower depths, (2) emplacement of the sulphides within a pre-existing pillow mound or breccia pile which has now been largely replaced leaving behind relics of the original pillow talus, (3) partial burial of the deposit by lava flows at intermediate stages during its growth, or (4) disruption of the breccias and footwall rocks by faulting within the mound. The construction of the TAG active mound therefore is interpreted to be substantially a process of hydrothermal replacement and mineralization in the upflow zone coupled with mass wasting,

brecciation and cementation of material that was precipitated on the sea floor.

The abundance of anhydrite preserved within the mound (estimated to be $\sim 10^5 \text{ m}^3$) was unexpected. Owing to its retrograde solubility (that is, its solubility decreases with increasing temperature), the precipitation of anhydrite from sea water at a water depth of 3,650 m requires the sea water to be heated to $>150^\circ\text{C}$. Hence the presence of anhydrite reflects the high temperatures maintained within the mound, and its abundance implies substantial entrainment of cold sea water into the interior of the deposit. Estimates⁵ of the convective heat flux from the Black Smoker complex are in excess of 200 MW. The focusing of this flux in a small area suggests that the vigorous high-velocity discharge of hot, hydrothermal fluids results in sea water being drawn into the edges of the mound and into the upflow zone beneath the Black Smoker complex¹² where it is heated by mixing with the high-temperature hydrothermal fluids. By this mechanism, many cubic kilometres of sea water may have been entrained over the life of the hydrothermal system. During long periods of inactivity, temperatures within the mound undoubtedly decreased below 150°C causing abundant dissolution of anhydrite and resulting in the formation of open spaces and collapse breccias.

The internal structure of the TAG mound bears a striking resemblance to ancient volcanic-hosted massive sulphide deposits preserved in ophiolites. Such deposits found in the ophiolites of Cyprus, Oman and Newfoundland are regarded as the closest ancient analogues of massive deposits currently forming at mid-ocean ridges. Sulphide and wall-rock breccias are well documented in Cyprus-type massive sulphides and include coarse sulphide conglomerates, quartz-cemented breccias, and chloritized and silicified basalt breccias^{16,19,20,21}. Metal-enriched zones similar to those in the upper few metres of the TAG active mound have been observed close to the palaeo sulphide-seawater inter-

FIG. 2 Diagram of the TAG active mound showing the surface morphology and distribution of venting, as well as the generalized and simplified internal structure based on the drilling results. Letters in brackets refer to the drillhole designations at each site, as shown in Fig. 1. Recovery of rock core ranged from <1 to 63% with an average of $\sim 12\%$. Four major zones within the mound and upflow zone can be distinguished. Zone 1. Massive pyrite breccias—clasts (up to 5 cm) of massive, granular pyrite in a porous, sandy pyrite matrix with minor amounts of anhydrite (Fig. 3b). Pyrite typically comprises $>75 \text{ vol.}\%$ of the rock. Zone 2. Pyrite–anhydrite breccias down to about 30 m.b.s.f.—rounded pyrite clasts (0.5–2 cm in size) in a matrix of anhydrite (Fig. 3c). Pyrite comprises 50–75 vol.%, anhydrite $>25 \text{ vol.}\%$. Pyrite–silica–anhydrite breccias that extend to $\sim 45 \text{ m.b.s.f.}$ —clasts of siliceous pyritic material, quartz–pyrite aggregates, and granular pyrite cemented mainly by quartz with cross-cutting anhydrite veins. Pyrite comprises 50–75 vol.%, silica $>10 \text{ vol.}\%$, and anhydrite $>10 \text{ vol.}\%$. Anhydrite veining is extremely well developed throughout this zone (Fig. 3d). Zone 3. Pyrite–silica breccias—large ($\leq 10 \text{ cm}$) grey fragments of siliceous material (pre-existing mineralized and silicified wall rock) and quartz–pyrite clasts in a matrix of fine-grained quartz (Fig. 3e). Pyrite comprises $>50 \text{ vol.}\%$, quartz matrix $>10 \text{ vol.}\%$. Silicified wall-rock breccias—clasts of basaltic fragments (1–5 cm in size) that are recrystallized to quartz, pyrite and a phyllosilicate, although some contain relict igneous textures. Zone 4. Chloritized basalt breccias occur below $\sim 100 \text{ m}$ —chloritized and weakly mineralized basalt fragments (1–5 cm in size) cemented by quartz and pyrite, and cross-cut by veins of pyrite, quartz and quartz + pyrite (Fig. 3f). (This figure was drawn by E. P. Oberlander.)

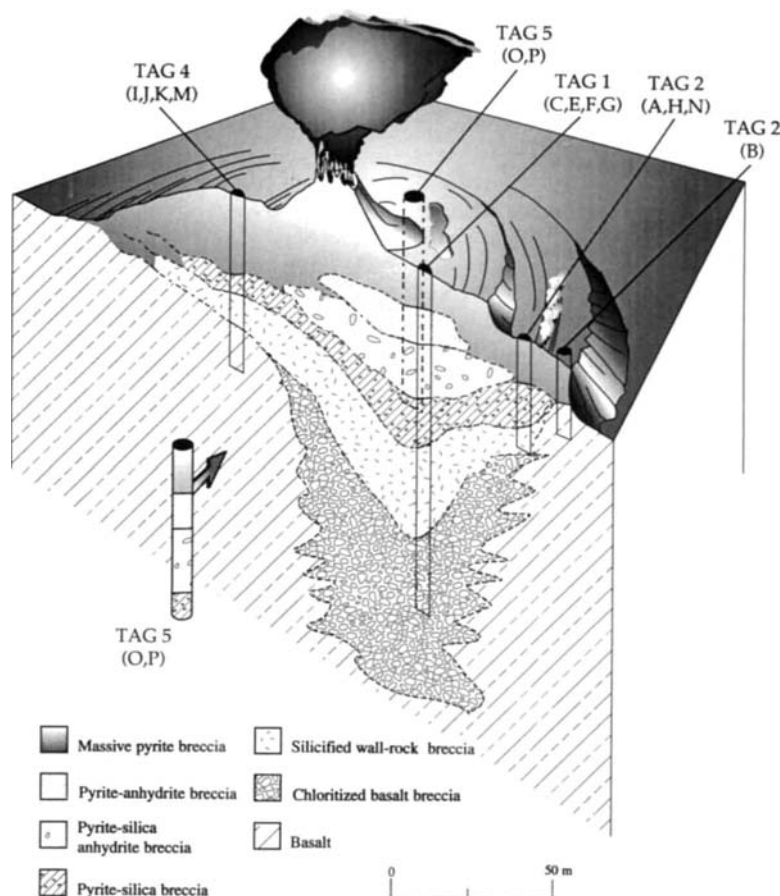
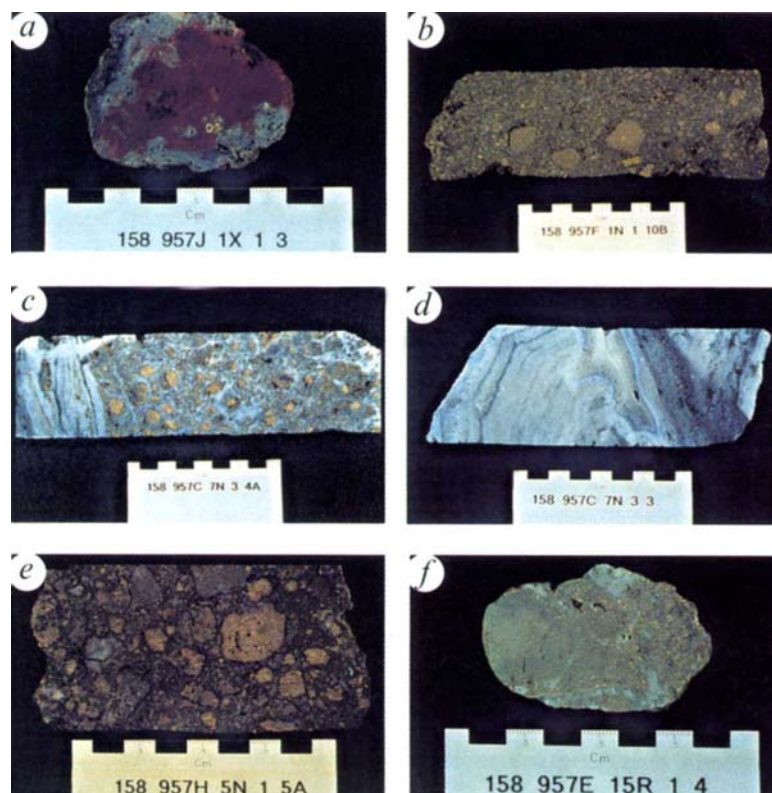


FIG. 3 Examples of different rock types comprising the sub-surface portion of the TAG active hydrothermal mound. Scale bar in each photograph is divided in 1 cm intervals. **a**, Red and grey siliceous material (sample 158-957J-1X-1, piece 3) from 0.2 m.b.s.f. from the TAG-4 area. It consists of massive iron oxide bearing silica (red colouration) surrounded by grey quartz and chalcedony. The contacts between red and grey siliceous zones are irregular, sharp or gradational. Pyrite is rare in the red silica, but occurs as fine-grained, disseminated euhedral grains and as colloform layers in the grey silica. **b**, Pyrite breccia (sample 158-957F-1N-1, piece 10B from 1.4 m.b.s.f. from the TAG-1 area) composed of clasts of massive granular pyrite in a porous, sandy pyrite matrix with minor anhydrite (<10 vol.%). An angular chalcopryite clast, similar in texture to the innermost, monomineralic chalcopryite lining the black-smoker chimneys¹² is present. **c**, Pyrite-anhydrite breccia (sample 158-957C-7N-3, piece 4A from 23 m.b.s.f. from the TAG-1 area) is the principal sulphide type in the anhydrite-rich zone. It consists of rounded clasts of massive granular pyrite in a matrix of massive to semimassive anhydrite with minor disseminated sulphides. With increasing depth, siliceous clasts and quartz-pyrite clasts become increasingly common as this rock type grades into a siliceous pyrite-anhydrite breccia. The contact with the overlying zone of anhydrite veins (which is at least 30 cm thick) can be seen. This portion of the vein contains at least two, and possibly three, generations of anhydrite veins. **d**, Portion of the same anhydrite vein (sample 158-957C-7N-3, piece 3) shown in **c** and oriented in the same direction. It consists of massive granular banded anhydrite with fine, disseminated pyrite and/or chalcopryite along the growth bands. The 2-cm-wide, dark-grey band comprises a second vein type that occurs locally within the large anhydrite vein. It is composed of coarsely crystalline anhydrite with abundant, uniformly disseminated pyrite and chalcopryite. Late coarse-grained, barren anhydrite replaces earlier massive banded anhydrite as concentric zones around a late fracture. **e**, Pyrite-silica breccia (sample 158-957H-5N-1 piece 5A from 27 m.b.s.f. from the TAG-2 area) composed of rounded clasts of pyrite in a matrix of very fine-grained quartz. Clasts of siliceous wall rock (consisting of grey-to-white quartz plus variable amounts of pyrite) are also present. **f**, Chloritized basalt breccia



(sample 158-957E-14R-1, piece 10 from 102 m.b.s.f. from the TAG-1 area) consisting of grey basaltic fragments pervasively altered to chlorite, a clay mineral (perhaps sericite), and/or quartz in a matrix of dark-grey quartz and pyrite. Pyrite occurs as 1–4 mm wide zones and chalcopryite occurs as rims, on the margins of altered basaltic clasts. Veins of pyrite, quartz and pyrite+quartz cross-cut the basaltic fragments.

face in volcanogenic massive sulphide deposits on land, and the siliceous material is analogous to the red cherts (jasper) associated with many of the Cyprus-type deposits¹⁶. The abundance of breccia ores and sulphide conglomerates in the Cyprus deposits was previously explained as post-depositional weathering and destruction of the mounds^{20,22}. The observations at TAG suggest, however, that these breccias were probably formed at the time of high-temperature venting and were initially cemented by

anhydrite. Anhydrite is now absent from the Cyprus deposits because it dissolved soon after high-temperature hydrothermal activity had ceased. Although the complex assemblage of breccias partly confirms models of deposit growth deduced from the study of Cyprus-type deposits, the importance of anhydrite in this process had not been predicted, and would not have been revealed without drilling an active sea-floor massive sulphide deposit. □

Received 9 June; accepted 16 September 1995.

1. Spooner, E. T. C. in *Deep Drilling Results in the Atlantic Ocean: Ocean Crust* (eds Talwani, M., Harrison, C. G. & Hayes, D. E.) 429–431 (Am. Geophys. Un., Washington DC, 1978).
2. Spooner, E. T. C. *Geol. Ass. Canada Spec. Pap.* **20**, 685–704 (1980).
3. Rona, P. A. & Scott, S. D. *Econ. Geol.* **88**, 1935–1975 (1993).
4. Herzig, P. M. & Hannington, M. D. *Ore Geol. Rev.* (in the press).
5. Rona, P. A. et al. *Econ. Geol.* **18**, 1989–2017 (1993).
6. Rona, P. A. et al. *J. geophys. Res.* **98**, 9715–9730 (1993).
7. Rona, P. A., Klinkhammer, G., Nelsen, T. A., Trefry, J. H. & Elderfield, H. *Nature* **321**, 33–37 (1986).
8. Lalou, C. et al. *Earth planet. Sci. Lett.* **97**, 113–128 (1990).
9. Lalou, C. et al. *J. geophys. Res.* **98**, 9705–9713 (1993).
10. Humphris, S. E., Kleinrock, M. C. & Deep-TAG Team (abstr.) *Eos* **75**, 660 (1994).
11. Thompson, G., Humphris, S. E., Schroeder, B., Sulanowska, M. & Rona, P. A. *Can. Mineralogist* **26**, 697–711 (1988).

12. Tivey, M. K., Humphris, S. E., Thompson, G., Hannington, M. D. & Rona, P. A. *J. geophys. Res.* **100**, 12527–12555 (1995).
 13. Campbell, A. C. et al. *Nature* **335**, 514–519 (1988).
 14. Edmond, J. M., Campbell, A. C., Palmer, M. R. & German, C. R. (abstr.) *Eos* **71**, 1650–1651 (1990).
 15. Edmond, J. M. et al. in *Hydrothermal Vents and Processes* (eds Parson, L. M., Walker, C. L. & Dixon, D. R.) 77–86 (Geol. Soc. Spec. Publ., London, 1995).
 16. Franklin, J. M., Lydon, J. W. & Sangster, D. F. *Econ. Geol.* **75**, 485–627 (1981).
 17. Strens, M. R. & Cann, J. R. *Tectonophysics* **122**, 307–324 (1986).
 18. Strens, M. R. & Cann, J. R. *Geophys. J. R. astr. Soc.* **71**, 225–240 (1982).
 19. Adamides, N. G. thesis, Univ. Leicester (1984).
 20. Constantinou, G. in *Proc. Int. Ophiolite Symp. on Ophiolites* (ed. Panayiotou, A.) 663–674 (Cyprus Geol. Surv. Dept., Nicosia, 1980).
 21. Lydon, J. W. *Geol. Surv. Can. Pap.* **84-1A**, 601–610 (1984).
 22. Constantinou, G. *Geol. Ass. Can. Spec. Pap.* **14**, 187–210 (1976).
- ACKNOWLEDGEMENTS. We wish to express our gratitude to the drilling, ship and scientific staff on board the D/V JOIDES Resolution for their help during ODP Leg 158.

Rapid degradation of atmospheric methyl bromide in soils

J. H. Shorter*, C. E. Kolb*, P. M. Crill†, R. A. Kerwin†, R. W. Talbot†, M. E. Hines† & R. C. Harriss†

* Center for Chemical and Environmental Physics, Aerodyne Research Inc., Billerica, Massachusetts 01821, USA

† Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, New Hampshire 03824, USA

METHYL bromide (CH_3Br), a widely used agricultural fumigant, may be an important source of atmospheric bromine radicals, which destroy stratospheric ozone^{1–10}. In current models of this compound's atmospheric behaviour, the main sinks are taken to be oxidation by hydroxyl radicals, photolysis and uptake by the oceans^{1–5, 7, 9, 10}. But there is also evidence that CH_3Br is consumed in soils^{8, 11–13}. Here we report laboratory and field experiments which show that, when exposed to a variety of soil types at low mixing ratios, CH_3Br is rapidly and irreversibly removed to below the levels found in the global atmosphere. We show that the uptake process is bacterially mediated. We estimate the global annual soil sink to be 42 ± 32 Gg; coupled with other removal mechanisms, this suggests an atmospheric lifetime for CH_3Br of about 0.8 yr, just half the previous best estimate⁹, and an ozone depletion potential that is about 30% smaller than the previous estimate¹.

Methyl bromide is the most abundant gaseous bromine species in the atmosphere, with an average atmospheric mixing ratio in the Northern Hemisphere of 10–15 parts per trillion by volume (p.p.t.v.), with values in the Southern Hemisphere 15–30% lower^{1–5}. Important atmospheric sources for CH_3Br include: biological production and emission from marine environments^{1–5}, biomass burning^{1–6}, exhaust from cars burning leaded petrol¹ and atmospheric emission of synthetic CH_3Br following its use as a soil, commodity or structural fumigant^{1–7}. Soil fumigation accounts for ~80% of synthetic CH_3Br use, and some soil pests and pest complexes cannot currently be controlled without it⁷. The economic viability of a wide variety of crops is threatened if agricultural uses of CH_3Br are restricted before adequate substitutes are available and certified for crop use^{7, 8}. The principal loss process for atmospheric CH_3Br is reaction with photochemically produced hydroxyl radicals, which, coupled with a small photolysis sink, has led to previous estimates of atmospheric lifetime of between 1.7 and 2.0 years^{1–4, 9, 10}, although this will be revised downwards to 1.6–1.7 years with the recent re-evaluation of the global tropospheric OH radical concentration field¹⁴. The chemical destruction rate of CH_3Br in salt water is well characterized¹⁵, supporting an ocean uptake lifetime ranging between 3.7 and 2.7 years, with the lower estimates arising from the most recent evaluation^{5, 16, 17}. Less is known about land surface sinks for CH_3Br ; studies of the interaction of high concentrations, 10^{-10} – 10^{-4} parts per million by volume (p.p.m.v.), with soils have identified methylation of organic matter as a major chemical sink^{8, 11}. In addition, Oremland and co-workers¹² have shown that CH_3Br is destroyed by reaction with sulphide and chloride anions in anaerobic salt marsh soil. But these chemical sinks apparently saturate at the high exposure levels characteristic of soil fumigation, and much of the applied CH_3Br escapes to the atmosphere^{8, 11}. Recent observations by Oremland *et al.*¹³ indicate that aerobic methanotrophic bacteria can consume significant amounts of CH_3Br when exposed to gaseous mixing ratios between 10 and 1,000 p.p.m.v.

We have investigated soil consumption of CH_3Br at low mixing ratios (<10 to ~14,000 p.p.t.v.) in both laboratory and field experiments. The laboratory experiments included manipulation of soil temperature, soil moisture and headspace atmosphere, and the addition of antibiotics to the soil. Field experiments included measurements of CH_3Br uptake rates at a variety of sites and the measurement of the diffusion rates of an inert tracer gas.

The soils for laboratory experiments were collected at the following sites: temperate forest soil (UNH College Woods, Durham, New Hampshire, USA), temperate agricultural soil (UNH Cornfield, Durham, New Hampshire, USA), tropical forest/agricultural soils (Costa Rica), and sandy boreal soils (Manitoba, Canada). Field experiments were located at the temperate soil collection sites and an additional site was located in a temperate grassy lawn (Lee, New Hampshire, USA). All experiments included analysis of the soils for gravimetric moisture content, pH, organic matter content and *in situ* bulk density.

Laboratory CH_3Br soil uptake rates were measured by placing 10 g of soil into 150 ml Pierce vials that were sealed with neoprene septa and aluminium seals then suspended in a controlled-temperature water bath. Known quantities of CH_3Br were injected into the vials, yielding headspace mixing ratios of 5–14 parts per billion by volume (p.p.b.v.). Duplicate vials were sampled sequentially by sweeping out the vial headspace with nitrogen at 100 ml min⁻¹ and cryotrapping onto a Poropak Q and quartz-wool plug in a Teflon loop suspended in a dry ice/isopropanol cryotrap (~–75 °C). The samples were then analysed using a gas chromatograph equipped with an electron capture detector (GC-ECD)²⁰.

A variety of soil types were incubated in the laboratory to determine bulk soil uptake activity coefficients for low mixing ratios of CH_3Br (Table 1). Initial headspace mixing ratios of 5–14 p.p.b.v. were still much higher than ambient levels, but were 1,000 times smaller than those used in previous studies^{12, 13}. All surface soils tested consumed CH_3Br rapidly. Blank experiments, and experiments with autoclaved soil consumed no measurable CH_3Br , demonstrating that physisorption or chemisorption onto container surfaces, sampling system components, or soil surfaces was a negligible loss process. Temperate forest soils taken from 0–3 cm depth demonstrated the fastest uptake rates, drawing CH_3Br levels down to below the detection limit in 5–10 min. Activity declined with depth in the boreal and temperate forest soils (Fig. 1).

This contrasts with the depth distribution of CH_4 oxidation activity in these soils²¹. There was an increase in CH_3Br con-

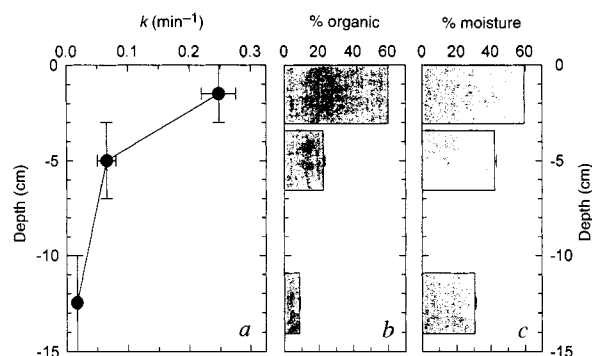


FIG. 1 Depth distribution of CH_3Br consumption in temperate forest soils. a, Initial depletion rates of CH_3Br (k) in soils from three depth intervals (vertical lines), incubated in the laboratory at 25 °C. b, Percentage dry weight loss on ignition at 435 °C of the soils (% organic) from each depth. c, Percentage weight change of the soil samples (% moisture) after drying to constant weight at 50 °C. In each panel, the horizontal lines are the standard deviation of five replicates at each depth.

† To whom correspondence should be addressed.

TABLE 1 Uptake of CH_3Br by different soil types

Soil type	k (min^{-1})	k (min^{-1} per g dry wt)	Soil moisture content (wt%)	Organic matter content (wt%)
Temperate forest				
Litter	0.183 ± 0.019	0.296 ± 0.0302	69	>95%
0–3 cm	1.14 ± 0.10	0.284 ± 0.028	54.8–64.4	56.7 ± 63.8
3–7 cm	0.376 ± 0.087	0.0652 ± 0.015	41.4–43.3	21.9–23.7
10–15 cm	0.117 ± 0.011	0.0168 ± 0.0016	29.9–31.3	8.2–9.2
Temperate agricultural surface soil	0.0805 ± 0.0421	0.00356 ± 0.0033	9.8–15.0	8.9
Sandy boreal forest surface soil	0.0894 ± 0.0002	0.0102 ± 0.0005	9.8–15.8	3.2–7.1
Tropical forest surface soil	0.0415 ± 0.0114	0.00652 ± 0.0011	35.0–40.8	21.7–27.6

* All experiments were performed at 25 °C, with an initial MeBr headspace mixing ratio of 13.5 p.p.b.v. Bulk soil uptake rate constants, k , are given as observed and as normalized to soil dry weight.

sumption of $18.1 \pm 11.2\%$ (standard error) by 0–3 cm and $24.2 \pm 7.4\%$ by 3–7 cm soils in the presence of 3–5% CH_4 . Normally a decrease would be expected if CH_4 oxidizing bacteria were also consuming CH_3Br . Figure 1 also illustrates the general relationship of decreasing CH_3Br uptake activity with decreasing moisture and organic matter content.

The uptake of CH_3Br seems to be an aerobic prokaryotic process. Consumption is stopped by autoclaving, and the temperature response is consistent with biological consumption. Maximum rates of uptake are between 25 and 37 °C. Activity is inhibited up to 98% by incubating soils under an anaerobic, nitrogen, headspace.

Antibiotics were used as a rough indicator of possible CH_3Br uptake activity by fungi or other higher organisms in surface soils. There is no change in activity in response to cycloheximide additions which would inhibit many non-bacterial organisms. The treatment of soils with the bactericides tetracycline and/or chloramphenicol lowered consumption rates of CH_3Br by 80% or more, confirming that eukaryotic or other cycloheximide-sensitive microflora were not consuming CH_3Br .

Field measurements of CH_3Br were made with a cubic 30-l Teflon chamber placed over the soil on a Teflon-coated aluminium collar. The collar was installed in the soil to a depth of 15 cm with 2.5 cm above the ground surface. A fan was used to circulate the chamber air. An aliquot of CH_3Br standard was injected into the chamber yielding an initial headspace mixing ratio of 600 p.p.t.v. The depletion of CH_3Br in the chamber headspace was measured by collecting a series of samples over 10–20 minutes. The samples were returned to the laboratory and CH_3Br was analysed using a GC-ECD after cryotrapping²⁰.

Chamber flux measurements were made at three locations: forest, grassy lawn and cornfield sites, with each site supporting 10–13 individual flux measurements. The 600 p.p.t.v. CH_3Br added to the chamber headspace was often consumed to undetectable limits (~ 2 p.p.t.v.) within 10 minutes. A chemically

inert tracer (SF_6) was often added to the headspace of the chamber to quantify the diffusive soil loss of the added gases. No measurable CH_3Br or SF_6 uptake on the field chamber walls was observed in controlled laboratory tests. Average uptake flux values and standard deviations, normalized to ambient CH_3Br levels, are shown in Table 2 for these sites. The decline of CH_3Br was always faster than that predicted from the SF_6 loss rate, except in the case of the cornfield soil where the loss was equal to the predicted diffusive loss rate (Fig. 2). This is direct evidence that these soils are actively consuming CH_3Br at chamber headspace mixing ratios near and below those found in the ambient atmosphere. The cornfield soils tested were very dry (10–15%) and poor in organic matter (9%). The low bulk soil uptake activities measured by incubation of cornfield soils taken from 0–3 cm depth are reflected in the low flux rates. Diffusion of CH_3Br into the soil appears to be the rate-limiting step at the cornfield sites.

Although the measurements reported here demonstrate a significant soil sink for CH_3Br , its effect on the compound's atmospheric lifetime and ozone depletion potential is not easy to quantify, requiring an estimate of the globally integrated annual soil sink for CH_3Br . Such an estimate can be generated by assigning daily loss flux estimates (based on the kinetic data reported here) to various biomes, then multiplying by biome areas and season lengths. The latter adjustment recognizes that active microbial uptake is repressed in temperate and boreal regions when the soil is frozen. Details of this estimate are also given in Table 2, where biome classifications and corresponding areas are adopted from a similar estimate of methane soil consumption rates by Born *et al.*²². Because both our work and that of Orem-land *et al.*¹³ indicate that bacteria which consume CH_3Br thrive in warm aerated soils, and are similar to the methanotropes responsible for soil methane consumption, it seems reasonable to adopt the same biomes assignment used by Born *et al.* to calculate the magnitude of the atmospheric methane soil sink

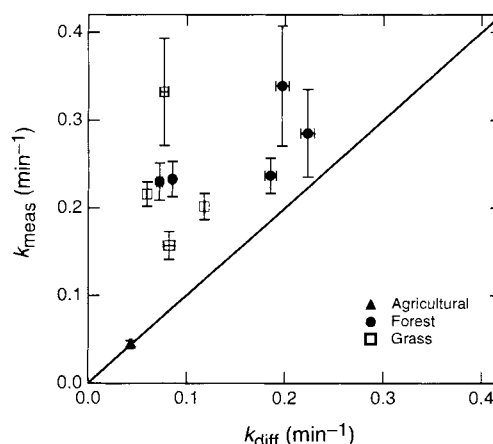


FIG. 2 Measured consumption plotted against diffusion uptake rates of CH_3Br by surface soils. Initial rates of consumption of CH_3Br (k_{meas}) were determined in the field by measuring the change in CH_3Br mixing ratios over time in a chamber headspace. Diffusion uptake rates (k_{diff}) are predicted from the diffusive loss of the inert tracer, SF_6 , over the same time period from the same chamber as in the CH_3Br loss rate determination. If the loss of CH_3Br from the chambers is purely diffusive, then the symbols would fall on the solid 1:1 line.

TABLE 2 Annual global soil uptake estimate for CH₃Br

Biome	Area (m ²)	Active season (d yr ⁻¹)	k _{obs} [*] (min ⁻¹)	Flux (g m ⁻² d ⁻¹)	Total (10 ⁹ g yr ⁻¹)
Tropical forest and savanna	22.5 × 10 ¹²	365	0.0415 ± 0.0114	7.9 (± 0.18) × 10 ⁻⁷	6.5
Temperate forest and woodland shrubland	20.0 × 10 ¹²	240	0.245 ± 0.134	4.5 (± 2.5) × 10 ⁻⁶	21.7
Temperate grassland	9.0 × 10 ¹²	240	0.242 ± 0.079	4.5 (± 1.5) × 10 ⁻⁶	9.7
Boreal forest	12.0 × 10 ¹²	180	0.0894 ± 0.0002	7.9 (± 0.18) × 10 ⁻⁷	1.7
Cultivated land	14.0 × 10 ¹²	240	0.0805 ± 0.0421	7.9 (± 0.18) × 10 ⁻⁷	2.7
				Total	42

* k_{obs} is the observed first-order rate constant for methyl bromide uptake into soils. For the soil types of temperate forest and woodland/shrubland and cultivated land, k_{obs} is the rate constant determined by our field measurements. For the other soil types, k_{obs} is the rate constant of the surface soil layer as measured in laboratory incubations, and the stated error limits represent one standard deviation for all measured samples which varied between 2 and 5.

for our estimate of the CH₃Br soil sink. In Table 2 the average measured field chamber loss flux values (following ref. 16, these are normalized to an average CH₃Br mixing ratio of 11 p.p.t.v. which seems appropriate considering the higher levels found in the land-rich Northern Hemisphere) for our temperate forest sites and temperate grassland sites are assigned to the temperate forest/woodland/shrubland and temperate grassland biomes, respectively. Field observations showed that these (nearly identical) loss fluxes were not limited by soil diffusion. Consumption by temperate agricultural soils, however, was apparently diffusion-limited. Because laboratory studies showed that soil uptake kinetic rates (also shown in Table 2) for our temperate agricultural, sandy boreal, and tropical forest soils, while still rapid and irreversible, were all a factor of 3 to 6 lower than those for the temperate forest and temperate grassland surface soils, these biomes were assigned uptake flux values characteristic of soil-diffusion-controlled kinetics. The diffusion flux value is based on field measurements of mass-corrected soil-diffusion-controlled uptake using SF₆ as an unreactive tracer. This latter assignment is justified by the fact that all of our laboratory work on unmodified soils showed that CH₃Br uptake was rapid and irreversible. Note that the largest normalized CH₃Br loss flux used in Table 2 corresponds to a deposition velocity of 0.12 cm s⁻¹, well below the maximum value measured for many reactive atmospheric species.

The total estimated annual soil sink for CH₃Br from the data in Table 2 is 42 Gg yr⁻¹. This can be regarded as a reasonably conservative estimate; first, because we assume no uptake in more anaerobic swamp, bog, fen and tundra biomes and second, because the extents of the aerated temperate soil biomes chosen by Born *et al.*²² are moderate. If we assign all of the temperate and tropical grassland, shrubland, woodland and forest areas identified in the 1° × 1° (latitude × longitude) analysis of satellite imagery analysed by Matthews²³, the extent of the forest/woodland/shrubland grouping increases by nearly 50% and the grassland by a factor of 3, with modest increases in the area of both boreal forest and cultivated land. The net result of these expanded biome area assignments is to increase the estimated

soil uptake to 74 Gg yr⁻¹. It will clearly be necessary to determine seasonal soil sink flux terms at many more sites in order to provide a statistically robust global soil-sink loss rate. But, after assessing the variability in our kinetic data and its dependence on soil moisture, soil organic content and temperature, along with the uncertainty in areal assignments for potential soil sink biomes noted above, we estimate that the value calculated in Table 2 is valid to ±75% or 42 (±32) Gg yr⁻¹.

An additional CH₃Br loss term of this magnitude would have a significant effect on its atmospheric lifetime. Given a total atmospheric burden of CH₃Br of ~143 Gg (ref. 17), a soil loss term of 42 Gg yr⁻¹ would correspond to a soil-loss lifetime of ~3.4 years.

The total atmospheric lifetime, τ_{total}, can be determined from the sum of the reciprocal lifetimes due to each loss process²⁴:

$$\frac{1}{\tau_{\text{total}}} = \frac{1}{\tau_{\text{photochem}}} + \frac{1}{\tau_{\text{ocean}}} + \frac{1}{\tau_{\text{soil}}}$$

From the currently published 'best estimate' values of 1.7 years for the photochemical lifetime⁹, 2.7 years for the ocean uptake lifetime^{3,17}, and 3.4 years for the soil uptake lifetime, we can estimate a total atmospheric lifetime of 0.8 years, or about half that attributed to the photochemical sink alone. This would reduce the calculated WMO 'best estimate' ozone depletion potential for CH₃Br by one-third, from 0.6 (based on a 1.3-year lifetime)¹ to 0.4.

Although the photochemical- and ocean-loss terms for CH₃Br are relatively well known, a more precise determination of the soil sink reported here will require significant additional effort. Furthermore, because of the greater land mass in the Northern Hemisphere, the additional CH₃Br soil sink identified here may require an additional northern hemispheric CH₃Br source in order to preserve the observed interhemispheric gradient in its atmospheric mixing ratio. The balance between natural and anthropogenic influences on the atmospheric concentration of CH₃Br is still uncertain. The real effect of the agricultural use of synthetic CH₃Br depends on the emitted fraction which reaches the stratosphere; an effective global soil sink significantly reduces that fraction. □

Received 20 June; accepted 25 September 1995.

- Ennis, G. A. (ed.) *Scientific assessment of ozone depletion: 1994* Ch. 10 (Rep. no. 37, V.2, Global Ozone Research and Monitoring Project, WMO, Geneva, Switzerland, 1995).
- Khalil, M. A. K., Rasmussen, R. A. & Gunawardena, R. J. *Geophys. Res. Lett.* **98**, 2887–2896 (1993).
- Reeves, C. E. & Penkett, S. A. *Geophys. Res. Lett.* **20**, 1563–1566 (1993).
- Singh, H., B. & Kanakidou, M. *Geophys. Res. Lett.* **20**, 133–136 (1993).
- Robert, J. M. *et al. Science* **267**, 1002–1005 (1995).
- Manó, S. & Andreae, M. O. *Science* **263**, 1255–1257 (1994).
- Methyl Bromide: Its Atmospheric Science, Technology and Economics* (United Nations Environmental Program, Nairobi, Kenya, 1992).
- Gan, J. *et al. Chemosphere* **29**, 2685–2700 (1994).
- Mellouki, A. *et al. Geophys. Res. Lett.* **19**, 2059–2062 (1992).
- Zhang, Z., Saini, R. D., Kurylo, M. J. & Huie, R. E. *Geophys. Res. Lett.* **19**, 2413–2416 (1992).
- Arvieu, J. C. *Acta Horticult.* **152**, 267–287 (1993).

- Oremland, R. S., Miller, L. G. & Strohmaier, F. E. *Envir. Sci. Technol.* **28**, 514–520 (1994).
- Oremland, R. S., Miller, L. G., Culbertson, C. W., Connell, T. L. & Jahnke, L. *Appl. envir. Microbiol.* **60**, 3640–3646 (1994).
- Prinn, R. G. *et al. Science* **269**, 187–192 (1995).
- Elliot, S. & Rowland, F. S. *Geophys. Res. Lett.* **19**, 1043–1046 (1992).
- Butler, J. H. *Geophys. Res. Lett.* **21**, 185–188 (1994).
- Yvon, S. A. & Butler, J. H. *Geophys. Res. Lett.* (in the press).
- Yagi, K., Williams, J., Wang, N. Y. & Cicerone, R. J. *Proc. natn. Acad. Sci. U.S.A.* **90**, 8420–8423 (1993).
- Yagi, K., Williams, J., Wang, N. Y. & Cicerone, R. J. *Science* **267**, 1979–1981 (1995).
- Kerwin, R. A. *et al. Anal. Chem.* (submitted).
- Crill, P. M. *Global Biogeochem. Cycles* **5**, 319–334 (1991).
- Born, M., Dörr, H. & Levin, I. *Tellus* **42B**, 2–8 (1990).
- Matthews, E. J. *Clim. Appl. Meteorol.* **22**, 474–487 (1983).
- Ravishankara, A. R. & Lovejoy, E. R. *J. chem. Soc. Faraday Trans.* **90**, 2159–2169 (1994).

ACKNOWLEDGEMENTS. We thank M. J. Kurylo, M. W. K. Ko, J. H. Butler and T. Duafala for comments on this manuscript. This work was supported by the Methyl Bromide Global Coalition.

A possible Early Cambrian chordate

J.-Y. Chen, J. Dzik*, G. D. Edgecombe†, L. Ramsköld‡ & G.-Q. Zhou

Nanjing Institute of Geology and Palaeontology, Chi-Ming-Ssu, Nanjing 210008, People's Republic of China

* Instytut Paleobiologii, PAN, Al. Zwirki i Wigury 93, PL 02-089

Warszawa, Poland

† Australian Museum, P.O. Box A285 Sydney South, New South Wales 2000, Australia

‡ Museum of Palaeontology, University of Uppsala, Norbyvägen 22, S-752 36 Uppsala, Sweden

THE first chordate recorded from the Early Cambrian is the cephalochordate *Yunnanozoon lividum* from the 525 million-year-old Chengjiang fauna. Chordate features of *Yunnanozoon* are a notochord and an expanded filter-feeding pharynx with an endostyle. Segmented musculature and metameric branchial arches are shared with cephalochordates and craniates. Metameric gonads and an anteriorly extended notochord indicate cephalochordate affinities. *Yunnanozoon* expands the range of cephalochordate morphology known from the younger *Pikaia gracilens* and crown group forms such as amphioxus. Our identification predicts that other chordate clades (tunicates and craniates) had evolved by the Late Atdabanian, in the main burst of the Cambrian Explosion.

Chordate interrelationships are traditionally¹ analysed primarily using developmental characters of extant taxa²: data from Early Palaeozoic fossils are often either regarded to be unhelpful³ or at best controversial. An alternative, the calcichordate hypothesis^{4,5}, relies on using Early Palaeozoic fossils for deducing chordate history. These forms, appearing in the Middle Cambrian, are usually regarded as echinoderms, however⁶. The only widely accepted Cambrian chordate, the Middle Cambrian *Pikaia*^{4,7,8}, awaits redescription. The contemporary but poorly known *Metaspriggina*⁹ is probably a chordate⁸, whereas alleged Early Cambrian tunicates^{10,11} are certainly misidentified. Here we interpret an Early Cambrian segmented organism of previously unknown affinities as a cephalochordate.

*Yunnanozoon lividum*¹², of the Early Cambrian (upper Atdabanian) Chengjiang fauna^{12,13}, was based on three specimens from Maotian Hill in Chengjiang, Yunnan, China. We have collected 22 further specimens. These are 25–40 mm long, usually flattened (Figs 1 and 2a) but occasionally showing relief (Fig. 2c, e), preserved as a thin, bluish grey to black film.

An anterior, pharyngeal region is followed by a dorsal, muscular region and a ventral, visceral region. The pharyngeal region includes upper, suprapharyngeal parts and lower, infrapharyngeal parts. These connect anteriorly through a widened, tunnel-shaped snout (Fig. 1a), open anteriorly. The pharynx is invariably mud-filled but the body never is. A distally tapering, light band, called notochord below, runs nearly straight from the snout tip through the suprapharyngeal tissues to the posterior body tip (Figs 1d and 2b, e). The notochord is just as flattened as other soft tissues, but shows weak relief when bent (Figs 1a and 2e).

Segmentally arranged structures above the notochord are in accordance with musculature being divided by myosepta into 22–24 myomeres (M1–M24) which continue to the posterior tip (Fig. 3d). Anterior to the long first myomere (M1) is an apparently unsegmented block (Figs 1a and 2). Individual myomeres may be shortened (Figs 1c and 2b, f), shortening increasingly away from the notochord. The first myoseptum is anteriorly concave (Fig. 1a and 2a, c); subsequent septa are nearly straight or weakly sigmoidal (amplified by oblique preservation), curving backwards dorsally and forwards ventrally, terminating against the upper notochord margin. The left and right body wall/

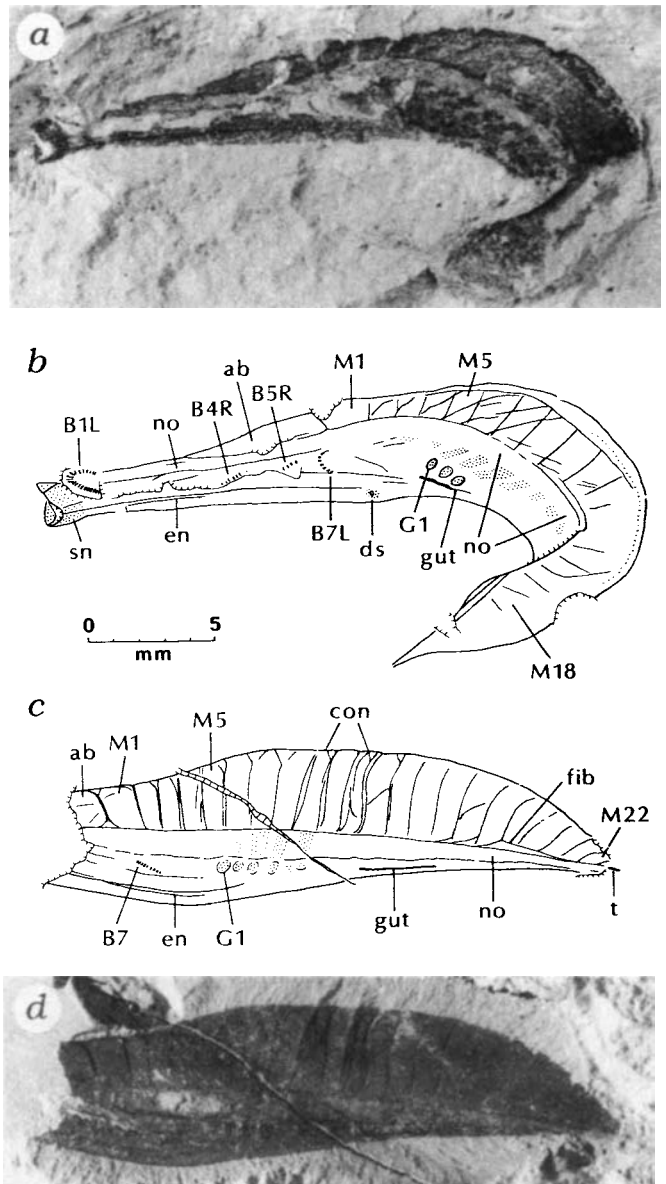


FIG. 1 *Yunnanozoon lividum* from Chengjiang. Abbreviations: B1L and B1R, left and right branchial rami (and so on); M1, myomeres; con, constricted myomeres, G1, gonads (and so on); sn, snout; en, endostyle; gut, tentatively identified gut; no, notochord; ab, anterodorsal block; ds, denticular structure; fib, notochord sheath fibres; t, tail tip. Scale bar applies to a–d. a and b, Photograph (before final snout preparation) and camera lucida drawing of strongly curved specimen (ELRC 52013a); drawing incorporates posterior body details from counterpart (ELRC 52013b). Note tunnel-like snout, and two sets of myoseptal lines between M1 and M7, one set transverse and the other strongly oblique. c and d, Camera lucida drawing and photograph of lateral specimen (ELRC 52011a). Note constricted M9 and M11, and the fibre bunches that appear to be partly torn loose (presumably during decay) from the notochord by M17/M18 septum, which may indicate that septa attached to a sheath of outer longitudinal fibres.

myoseptum junctions are occasionally seen together anteriorly (holotype and Fig. 3c) or beside shortened myomeres (Figs 1d and 2a).

Two parallel lines running longitudinally through the hypopharyngeal tissues from below M3–M4 to past the anteriormost branchial arch (Fig. 1a) may represent the endostyle (Fig. 3a, b). Seven evenly spaced, biramous structures are identifiable as branchial arches, B1–B7. B1 is set close to the snout. B7 terminates below M1. All rami run anterodorsally from relief-preserved points near the upper endostyle margin (Fig. 3b), B1–5, to the sides of the notochord, B6–7 to its lower surface (Fig. 2a, c).

† To whom correspondence should be addressed.

Each ramus consists of about 20 dark discs (Fig. 3b) and light, narrower interspaces, resembling the striped gill-bar mucocartilages in an ammocoete larva¹⁴. The rami were evidently elastic (Fig. 3b). A denticular, possibly mineralized structure (Figs 1a and 3a and type specimens¹²) is situated between B6 and B7 above the endostyle.

Below the notochord are up to 13 metamerically arranged, paired structures interpreted as gonads, the first beneath M5, becoming smaller posteriorly (Fig. 2d). Gonads show strong relief in three-dimensionally preserved specimens (Fig. 2c, e), retaining some relief also in flattened individuals (Fig. 1d). Gonads of the left and right sides alternate exsagittally (type specimens¹², and Figs 2c, e and 3c), indicating bilateral asymmetry. A straight (Figs 1d and 2e) or undulating (Figs 1a and 2a) dark line, shown on a longitudinally rotated specimen to be unpaired and positioned sagittally near the ventral margin (Fig. 2e), may represent the gut. It fades beneath M15, but no anus is discernible. Dorsal and ventral fins are indicated by narrow bands posteriorly (Fig. 3d).

Monophyly of Chordata is almost¹⁵ universally accepted^{2,3,5,16}. Among chordate synapomorphies^{2,3,5,17} potentially identifiable in fossils, *Yunnanozoon* possesses a notochord and endostyle. The rod in *Yunnanozoon* is likely to be a notochord because it: (1) is positioned dorsal to the pharynx as in euchordates¹⁸; (2) is often straight (Figs 1d and 2a), indicating stiffness; (3) resists longitudinal compression because shortened myomeres retain full length near the rod; (4) tapers at both ends and has a fibrous outer sheath; and (5) is consistently preserved, contrary to the inferred gut, in accord with decay resistance of a notochord sheath, as demonstrated for amphioxus¹⁹.

An expanded filter-feeding pharynx is synapomorphic for Chordata⁵ or a tunicate-cephalochordate clade²⁰. The pharynx of *Yunnanozoon*, positioned as in the cephalochordates, is identified by being filled with mud and having branchial rami attached ventrally to a band resembling an endostyle. The band is unlikely to be the ventral aorta, which should end posteriorly with an expanded heart. Regular spacing and consistent positions of the branchial rami relative to myomeres suggest metamerism, synapomorphic for cephalochordates and craniates¹⁸. The most compelling evidence for euchordate affinities, however, is segmented musculature. The marked shortening of compartmentalized dorsal units (Figs 1a and 2f), with concomitant indentation of the dorsal margin, can possibly be interpreted as post-mortem spasm in individual myomeres separated by myosepta.

An anteriorly extended notochord and metamerically repeated gonads indicate cephalochordate affinity (the former, however, is nearly approached by lamprey ammocoetes). The gonads in *Yunnanozoon* are metamerically arranged and positioned in the area occupied by the atrium in cephalochordates. Their apparent bilateral asymmetry in *Yunnanozoon* compares with the presence of gonads on the right side only in the extant cephalochordate *Asymmetron*, and their relative resistance to decay bears similarity with amphioxus¹⁹. Metameric organs in comparable position in the earliest lamprey are doubtfully gonads^{21,22}.

The protective atrium²³ and subdivided branchial bars of extant cephalochordates are evidently absent in *Yunnanozoon* and may be crown group specializations for ciliary feeding in a burrow. The elastic branchial rami of *Yunnanozoon* suggest an alternative strategy of pharyngeal contraction for pumping water, but there is little reason to doubt filter feeding. The small number of gill slits in *Yunnanozoon* compares with tunicates, amphioxus larvae and craniate ammocoetes, and is thus plesiomorphic. The ventral position of the notochord posteriorly and the concomitant extreme dorsalization of the myomeres are conspicuous in *Yunnanozoon*. The ubiquitous sublateral preservation indicates a laterally flattened body displaying considerable lateral flexure (Fig. 1a) but minor dorsoventral flexibility. This, along with myomere organization, indicates that *Yunnanozoon* undulated laterally like amphioxus. The denticular structure in *Yunnanozoon* is puzzling and has no known homologues in other

chordates, including conodonts^{24,25}, except possibly in *Conopiscius*³⁰.

Two other Cambrian taxa, *Pikaia*^{7,8} and *Emmonsaspis*^{26,27}, have been cited as cephalochordates. The late Early Cambrian *Emmonsaspis* is now recognized to be a frond-like organism²⁸. The Middle Cambrian *Pikaia* is plesiomorphic relative to *Yunnanozoon* and extant cephalochordates in apparently having the notochord truncated behind the 'head'²⁹ and lacking¹⁹ metamerically gonads. However, *Pikaia* resembles extant cephalochordates more than *Yunnanozoon* in its more dorsal notochord and abundant myomeres that reach the front of the body.

Note added in proof: Our attention has recently been drawn to the existence of an alternative possible interpretation of this species, and we understand that a manuscript on the subject has been submitted for publication. Although we feel this interpretation is unlikely for the reasons described above, further work will clearly be required to put the matter beyond doubt. □

Received 14 March; accepted 9 August 1995.

- Garstang, W. Q. *J. microsc. Sci.* **72**, 51–187 (1928).
- Schaeffer, B. *Evol. Biol.* **21**, 179–235 (1987).
- Maisey, J. G. *Cladistics* **2**, 201–256 (1986).
- Jefferies, R. P. S. *The Ancestry of the Vertebrates* (British Museum Natural History, London, 1986).
- Cripps, A. P. *Zool. J. Linn. Soc.* **100**, 27–71 (1990).
- Peterson, K. J. *Lethaia* **28**, 25–38 (1995).
- Conway Morris, S. (ed.) *Atlas of the Burgess Shale* (Palaeontological Association, London, 1982).
- Briggs, D. E. G., Erwin, D. H. & Collier, F. J. *The Fossils of the Burgess Shale* (Smithsonian Institution Press, Washington, 1994).
- Simonetta, A. M. & Insom, E. *Boll. Zool.* **60**, 97–107 (1993).
- Zhang, A. *Kexue Tongbao* **30**, 1643–1647 (1985).
- Zhang, A. *Scientia sin.* **B 30**, 888–896 (1987).
- Hou, X. G., Ramsköld, L. & Bergström, J. *Zool. Scripta* **20**, 395–411 (1991).
- Chen, J.-Y. & Erdtmann, B.-D. in *The Early Evolution of Metazoa and the Significance of Problematic Taxa* (eds Simonetta, A. M. & Conway Morris, S.) 57–76 (Cambridge Univ. Press, 1991).
- Dohrn, A. *Mitt. Zool. Stat. Neapel* **6**, 1–92 (1886).
- Løvtrup, S. *The Phylogeny of Vertebrata* (Wiley, London, 1977).
- Bleick, A. *Geobios* **25**, 101–113 (1992).
- Gans, C. *Biol. Rev.* **64**, 221–268 (1989).
- Jarvik, E. in *L'évolution dans sa Réalité et ses Diverses Modalités* 35–64 (Fondation Singer-Polignac & Masson, Paris, 1988).
- Briggs, D. E. G. & Kear, A. J. *Lethaia* **26**, 275–287 (1994).
- Jollie, M. *Zool. J. Linn. Soc.* **75**, 167–188 (1982).
- Janvier, P. & Lund, R. *vert. Paleont.* **2**, 407–413 (1983).
- Lund, R. & Janvier, P. *Geobios* **19**, 647–652 (1986).
- Young, J. Z. *The Life of Vertebrates* 3rd edn (Clarendon, Oxford, 1981).
- Aldridge, R. J., Briggs, D. E. G., Smith, M. P., Clarkson, E. N. K. & Clarke, N. D. L. *Phil. Trans. R. Soc. Lond. B* **340**, 405–421 (1993).
- Gabbott, S. E., Aldridge, R. J. & Theron, J. N. *Nature* **374**, 800–803 (1995).
- Resser, C. E. & Howell, B. F. *Bull. geol. Soc. Am.* **49**, 195–248 (1938).
- Termier, G. & Termier, H. *Evolution et Biocénose* (Masson, Paris, 1968).
- Conway Morris, S. *Palaeontology* **36**, 593–635 (1993).
- Gee, H. *Nature* **370**, 504–505 (1994).
- Briggs, D. E. G. & Clarkson, E. N. K. *Lethaia* **20**, 107 (1987).

ACKNOWLEDGEMENTS. We thank A. Ritchie, A. Seilacner, K. Fauchald, R. Strahan and D. Shu for discussions. This work was supported by the Chinese Academy of Science and the National Geographic Society. J. D. acknowledges support by the Polish Committee of Scientific Research. G. D. E. was supported by the Australian Museum Trust. L. R. was supported by the Wenner-Gren Foundation in China and by Magn. Bergvalls Stiftelse in producing the illustrations.

Limits to vertebrate locomotor energetics suggested by hummingbirds hovering in heliox

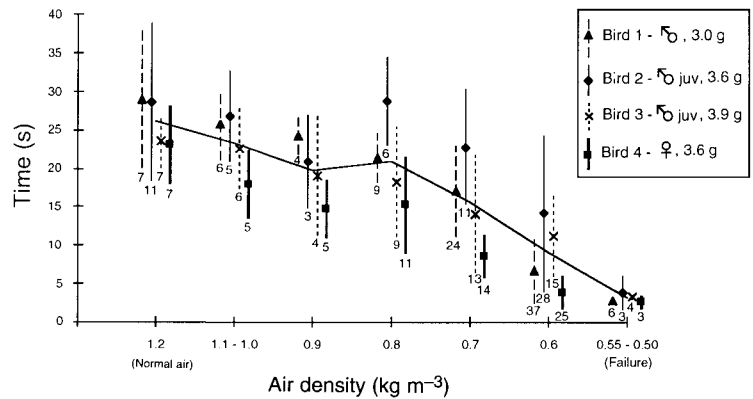
Peng Chai* & Robert Dudley*†

* Department of Zoology, University of Texas, Austin, Texas 78712, USA

† Smithsonian Tropical Research Institute, PO Box 2072, Balboa, Republic of Panama

OXYGEN consumption¹ and muscle power output² of hovering hummingbirds are among the highest recorded for vertebrates. Maximum performance of hummingbirds thus approaches the upper limits of vertebrate aerobic locomotion³. Because air density is a major determinant of aerodynamic power requirements⁴, hovering flight performances⁵ can be manipulated non-invasively using normoxic gas mixtures of variable density⁶. Here we show that limits

FIG. 1 Hover-feeding duration of four individual hummingbirds (mean $\pm$ s.d.; number of feeding events is indicated) in relation to air density reduction. Line links mean of individual means. Hover-feeding events were grouped by rounding air densities at which feeding events occurred and measurements were made. One adult male, two juvenile males and one female were used; all had intact flight feathers. Each bird was subjected to three density-reduction trials, one each for three consecutive days. Flight experiments were implemented within an airtight Plexiglas cube ($90 \times 90 \times 90$ cm), the contents of which were varied through gradual filling with 79% He/21% O₂. Nitrogen replacement by helium lasted 2–3 h until the bird failed while hover-feeding, at which point heliox filling was terminated, and reverse pumping of ambient air was initiated. This process was continued until density returned to approximately 0.8 kg m^{-3} , with hover-feeding duration lengthened to approximately 20 s. Aerodynamic failure was characterized by rapid descent from feeder to the chamber floor and subsequent inability to fly (air densities at failure: bird 1, $0.55 \pm 0.01 \text{ kg m}^{-3}$; bird 2, 0.50 ± 0.03 ; bird 3, 0.55 ± 0.04 ; and bird 4, 0.55 ± 0.02). Air density exhibited an exponential decline as the density difference between air and heliox was greatest at the onset of filling (air density was monitored acoustically as described previously⁶). The bird also reduced its hover-feeding duration and fed more frequently; more hover-feeding events were observed at lower air densities. Air temperature within the flight chamber averaged 25.3°C (range



$24.4\text{--}26.0^\circ\text{C}$); experiments were conducted at approximately sea level (160 m). All statistical tests used repeated-measures ANOVA modelling trial and density effects as within-subject sources of variation (data points were mean value of each density-trial combination). Results indicated a highly significant density effect ($P < 0.001$) but no trial effect. Further test of regression trends revealed significant linear ($P < 0.001$) and quadratic ($P < 0.05$) terms, arguing against a linearly declining slope.

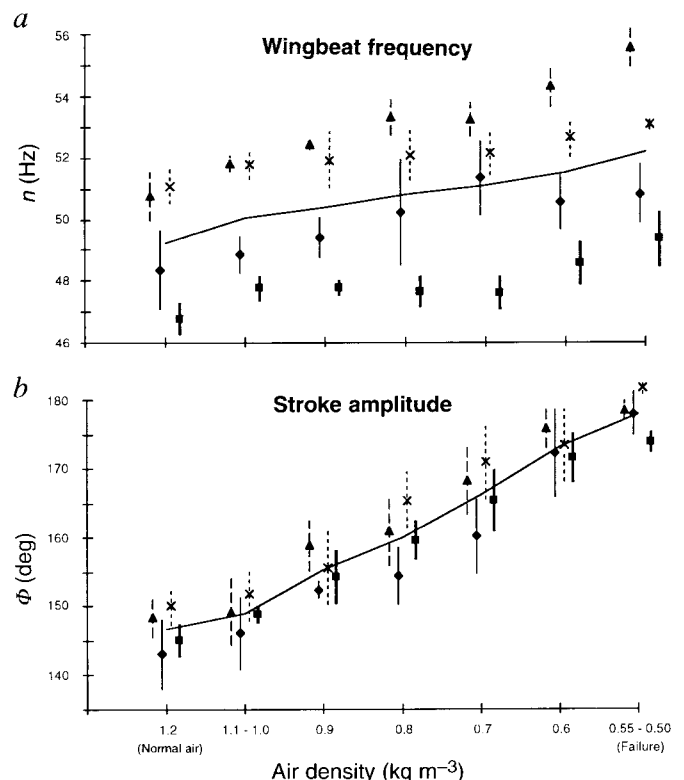
to the locomotor capacity of hovering ruby-throated hummingbirds are unequivocally indicated by aerodynamic failure at low densities less than half that of sea-level air. Hummingbirds demonstrate considerable power reserves, with muscle mass-specific power (assuming perfect elastic energy storage) averaging from 98 W kg^{-1} in normal air to a maximum value of 133 W kg^{-1} before aerodynamic failure. In contrast to such variable power expenditure, however, muscle efficiency remains approximately constant at 10%. Modulation of power output is attained primarily through variation in wing-stroke amplitude, with aerodynamic failure occurring near 180° .

By gradual replacement of sea-level air (density 1.2 kg m^{-3}) with heliox (79% He/21% O₂, density 0.4 kg m^{-3}), the density of the medium within which a bird hovers can be reduced. This reduction means that wingbeat kinematics must be altered to

match the increased requirements for mechanical power. Oxygen partial pressure is maintained at normal values, ensuring that diffusive limitations on metabolic power production cannot interfere with flight performance. We measured the mechanical performance, metabolic expenditure, and efficiency of hummingbird flight muscles during hovering in normal air and in gas mixtures of low densities. We know that maximum performance was unequivocally elicited in these experiments because the birds failed to sustain hovering.

Four ruby-throated hummingbirds (*Archilochus colubris*) were trained to feed while hovering at a mask that captured exhaled respiratory gases. Metabolic power input during hovering was obtained from measurements of oxygen consumption². Flight experiments were implemented within an airtight cube, the contents of which were varied through gradual filling with heliox.

FIG. 2 Wingbeat kinematics in relation to air density reduction (symbols and sample size as in Fig. 1). a, Wingbeat frequency; b, stroke amplitude. A video camera (Sony CCD-FX420 with a high-speed shutter of $1/4000$ s) was positioned vertically above the feeder to record horizontal projections of wingbeat kinematics at 60 fields s^{-1} . Wingbeat kinematics were determined from field-by-field video playbacks. Wingbeat kinematics measured for each hovering sequence included a, wingbeat frequency (n , derived over an interval of 2–10 s from the interaction frequency between wing motions and filming rate) and b, stroke amplitude (Φ). Mean kinematic variables were derived for each hovering sequence. Morphological parameters used in aerodynamic calculations were also determined on all filmed birds. These parameters included body mass (m), relative wing mass, wing length (R), total wing area (S , the area of both wing pairs), wing loading (mg/S , where g is acceleration due to gravity), and aspect ratio ($4R^2/S$). Non-dimensional parameters describing distributions of wing area and virtual mass were also derived. One bird (bird 1) was killed for *post mortem* analysis to determine wing mass and its distribution. Non-dimensional equivalents of wing mass parameters were used for the other three birds. Repeated-measures ANOVA was conducted on wingbeat frequency and stroke amplitude; results indicate a highly significant density effect for both variables ($P < 0.001$) but no trial effect.



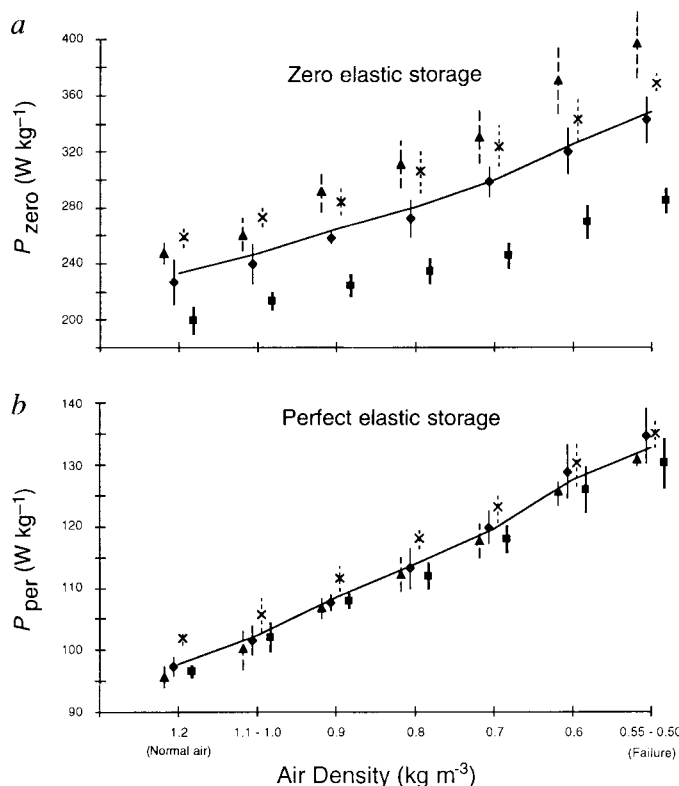


FIG. 3 Aerodynamic power production in relation to air density reduction (symbols and sample size as in Fig. 1). a, Zero elastic storage; b, perfect elastic storage. A detailed model of hovering aerodynamics⁷ was implemented using kinematic and morphological data obtained from individual hummingbirds. Because of approximately horizontal wing motions, downstroke and upstroke were assumed to contribute equally to vertical force production. For each hovering flight sequence, mechanical power requirements of flight were estimated by evaluating individual components of profile (P_{pro}), induced (P_{ind}) and inertial power (P_{acc}). P_{pro} represents energetic expenditure to overcome profile drag forces on the wings, P_{ind} corresponds to the power necessary to impart sufficient downwards momentum to the surrounding air so as to offset the body weight. The inertial power during the first half of a half-stroke was estimated from the moment of inertia of the wing and the maximum angular velocity assuming simple harmonic motion of the wings in the stroke plane⁵. Total inertial power requirements through the wingbeat will be zero if the kinetic energy of the oscillating wing mass and virtual mass can be stored elastically during deceleration of the wing stroke and subsequently released to reaccelerate the wings. Thus total power, $P_{per} = P_{pro} + P_{ind}$, assuming perfect elastic energy storage (b). However, assuming zero elastic energy storage (a), additional power will be required to accelerate the wing during the first half of a half-stroke. But power required for wing deceleration during the second half of the half-stroke is negligible, and aerodynamic power requirements can be met by kinetic energy of the decelerating wings⁷. Thus total power $P_{zero} = \frac{1}{2}(P_{pro} + P_{ind} + P_{acc})$. P_{per} and P_{zero} are expressed as muscle mass-specific form assuming flight-muscle proportion equals 25% of the body mass, as in published data for ruby-throated and other hummingbird species¹⁴. Repeated-measures ANOVA results indicate a highly significant density effect ($P < 0.001$) but no trial effect for both P_{per} and P_{zero} .

Wingbeat kinematics recorded at each hover-feeding sequence and morphological parameters of individual birds were used to estimate the mechanical power requirements of flight using contemporary aerodynamic models of hovering flight⁷. Muscle mass-specific power output for a flight sequence was calculated for the two cases of perfect and zero elastic storage of wing inertial energy, representing minimum and maximum estimates of required mechanical power, respectively. The mechanical efficiency of flight muscle was estimated as the ratio of power output to metabolic power input.

Initial density reduction resulted in a slow decline of hover-feeding duration (Fig. 1). Below a critical density of around 0.8 kg m^{-3} , the duration of feeding bouts exhibited a sharper decline. Further density reduction resulted in failure to sustain flight during a feeding bout lasting 2–4 s; the bird dramatically descended to the chamber floor, a behaviour never otherwise observed. Air densities at which hovering flight failed were less than half those of normal air, and were equivalent to hovering at altitudes of around 6000 m (average 0.54 kg m^{-3} , 65% $\text{He}/14\% \text{N}_2/21\% \text{O}_2$). Maximum performance was evaluated during the short interval before aerodynamic failure. Subsequent increases in chamber density through replacement of helium by nitrogen resulted in resumption of normal hovering behaviour, indicating that induced aerodynamic failure was fully reversible.

Reduced-density gas mixtures significantly altered wingbeat kinematics (Fig. 2). Both wingbeat frequency and stroke amplitude increased at lower densities, although the change in wingbeat frequency was small (4–10%). At failure densities, stroke amplitude increased by 20–24% and approached or slightly exceeded 180° . With reduced air density, muscle mass-specific power expenditure also increased under either assumption of zero or perfect elastic energy storage (increases of 42–61% (P_{zero}) and 32–38% (P_{per}) at failure; Fig. 3). Oxygen consumption rates at failure densities could not be reliably obtained owing to the short duration of hover-feeding. At air densities near 0.6 kg m^{-3} , however, maximum rates of oxygen consumption increased by 15–36% relative to normal hovering (Fig. 4). In contrast to such variable metabolic and power expenditure, muscle efficiency remained approximately constant at 10% (Fig. 4).

The steep decline in the duration of hover-feeding at low air densities has previously been observed using South American hummingbirds (*Amazilia fimbriata* and *Colibri coruscans*) in simulations of altitudinal change that also reduced oxygen partial pressure⁸. Because hummingbirds possess only aerobic flight muscles³, anaerobic metabolism with accompanying fatigue should play no role in these studies. Hovering energetics and muscle efficiency for ruby-throated hummingbirds hovering at sea level are similar to those of two North American *Selasphorus* species hovering at an elevation of 2,195 m (ref. 2). However, invasive weight-loading experiments with selected individuals of *Selasphorus* species averaged 117 W kg^{-1} muscle mass at maximum performance⁹, lower than the 133 W kg^{-1} obtained here. Similarly, *in vivo* strain measurements of maximum muscular power in pigeon flight estimated only 119 W kg^{-1} (ref. 10). Thus non-invasive measurements of maximum performance in hummingbirds yield values consistently higher than those previously obtained for birds. No mechanical performance data are available for bird muscles in isolated preparations to mimic *in vivo* muscle function, but the hawkmoth *Manduca sexta* is of comparable body mass to ruby-throated hummingbirds, and yielded a maximum of 130 W kg^{-1} (ref. 11). Other *in vitro* studies modelling power output during flight in insects, swimming in fishes, jumping in frogs, and running in lizards resulted in much lower maxima¹². Finally, the high maximum performance of hummingbirds indicates that a proposed upper sustainable limit of 100 W kg^{-1} for fast aerobic muscle does not apply, at least not for the operating frequencies and temperatures in the range considered here¹². The power output by ruby-throated hummingbirds was remarkably high considering that the actual structures of contraction (the myofibrils) account for only 50–60% of the muscle volume¹³.

Although ruby-throated hummingbirds could not hover in pure heliox, three species of orchid bees demonstrated this capability, with one species showing even higher levels of mechanical power output than the hummingbirds (160 W kg^{-1})⁶. For both the hummingbirds we studied and the orchid bees studied previously, the primary means of increasing power output was to increase muscle strain through an increase in stroke amplitude.

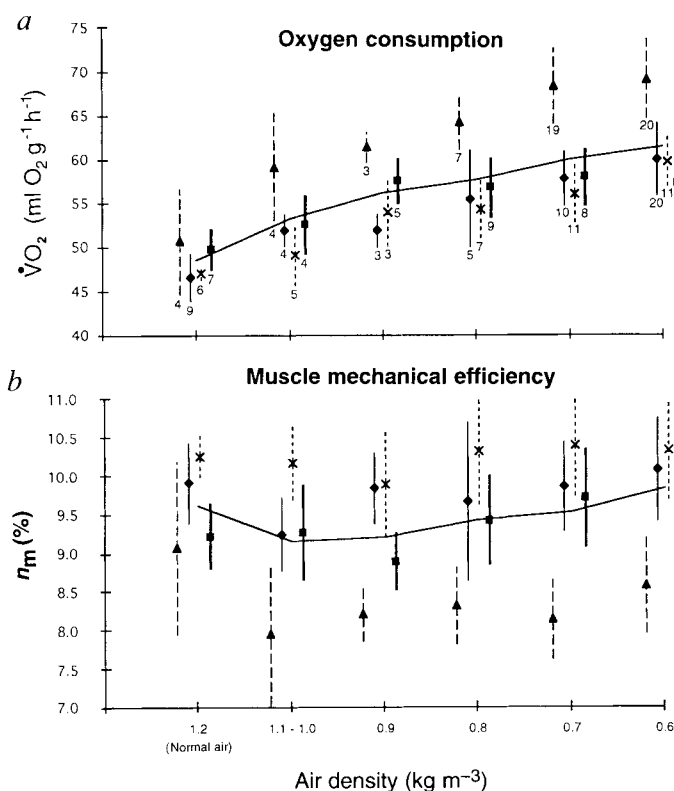


FIG. 4 a, Oxygen consumption and b, muscle mechanical efficiency in relation to air density reduction (symbols as in Fig. 1). Sample size is indicated and is reduced because only continuous traces of oxygen consumption longer than 5 s were used (hover-feeding events that were short or composed of intermittent feeding were not used). Oxygen consumption was determined using open flow mask respirometry, as described previously². Birds were trained to feed through a cylindrical mask. During hover-feeding, air was drawn through the feeder mask and was sampled and analysed by an Applied Electrochemistry S-3A/I Oxygen Analyzer. Metabolic power input (P_{input}) was determined from the rate of oxygen consumption by assuming a conversion factor of $21.1 \text{ J ml}^{-1} \text{ O}_2$ for carbohydrate utilization¹⁵ and a respiratory quotient of 1 (ref. 16). Mechanical flight muscle efficiency (n_m) was estimated as $P_{per}/(0.9 \times P_{input})$, assuming that 90% of total metabolism is attributable to the two pairs of major flight muscles¹⁷. P_{per} was used because hummingbirds can probably store kinetic energy elastically during the deceleration phase of the wing stroke². Calculated values of P_{zero} and associated muscle mechanical efficiency were implausibly high, for example, 349 W kg^{-1} of muscle with 25% mechanical efficiency at failure¹⁸. Repeated-measures ANOVA results indicate a highly significant density effect ($P < 0.001$) for oxygen consumption but not for muscle mechanical efficiency ($P = 0.078$). No trial effect was found for either variable.

Hummingbirds hovering in normal air already exhibit stroke amplitudes of $140\text{--}150^\circ$, and failure occurred universally near stroke amplitudes of 180° . By contrast, stroke amplitude for orchid bees hovering in heliox only reached 142° . Opposite wing interference and even impact is likely when stroke amplitude exceeds 180° . Thus limits for power production may be indicated by maximum stroke amplitude, a primary determinant of induced and total mechanical power during hovering flight. Further non-invasive studies of animal flight performance are necessary to evaluate such interactions between morphological design and the physiological limits inherent to skeletal muscle. □

- Norberg, U. M. *Vertebrate Flight* (Springer, New York, 1990).
- Weis-Fogh, T. *J. exp. Biol.* **56**, 79–104 (1972).
- Dudley, R. J. *exp. Biol.* **198**, 1065–1070 (1995).
- Ellington, C. P. *Phil. Trans. R. Soc. Lond. B* **305**, 145–181 (1984).
- Berger, M. J. *Ornithol.* **115**, 273–288 (1974).
- Wells, D. J. *J. exp. Biol.* **178**, 59–70 (1993).
- Dial, K. P. & Biewener, A. A. *J. exp. Biol.* **176**, 31–54 (1993).
- Stevenson, R. D. & Josephson, R. K. *J. exp. Biol.* **149**, 61–78 (1990).
- Josephson, R. K. *A. Rev. Physiol.* **55**, 527–546 (1993).
- Suarez, R. K., Lighton, J. R. B., Brown, G. S. & Mathieu-Costello, O. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4870–4873 (1991).
- Greenewalt, C. H. *Smithson. misc. Colln.* **144**, 1–46 (1962).
- Brobeck, J. R. & DuBois, A. B. in *Medical Physiology* Vol. 2 (ed. Mountcastle, V. B.) 1351–1365 (Mosby, St Louis, MO, 1980).
- Suarez, R. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 9207–9210 (1990).
- Lasiewski, R. C. *Physiol. Zool.* **36**, 122–140 (1963).
- Ellington, C. P. *J. exp. Biol.* **115**, 293–304 (1985).

ACKNOWLEDGEMENTS. We thank J. J. Bull, L. E. Gilbert, J. L. Larimer and M. J. Ryan for comments on the manuscript. This work was supported by an NIH NRSA and a University of Texas Reader Fellowship.

Bee foraging in uncertain environments using predictive hebbian learning

P. Read Montague*, Peter Dayan†, Christophe Person* & Terrence J. Sejnowski‡

* Division of Neuroscience, Baylor College of Medicine, Houston, Texas 77030, USA

† CBCL, Department of Brain and Cognitive Sciences, E25-201, MIT, Cambridge, Massachusetts 02139, USA

‡ Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037 and Department of Biology, University of California, San Diego, La Jolla, California 92093, USA

RECENT work has identified a neuron with widespread projections to odour processing regions of the honeybee brain whose activity represents the reward value of gustatory stimuli^{1,2}. We have constructed a model of bee foraging in uncertain environments based on this type of neuron and a predictive form of hebbian synaptic plasticity. The model uses visual input from a simulated three-dimensional world and accounts for a wide range of experiments on bee learning during foraging, including risk aversion. The predictive model shows how neuromodulatory influences can be used to bias actions and control synaptic plasticity in a way that goes beyond standard correlational mechanisms. Although several behavioural models of conditioning in bees have been proposed^{3–7}, this model is based on the neural substrate and was tested in a simulation of bee flight.

Real and colleagues^{8–11} performed a series of experiments on bumblebees foraging on artificial blue and yellow flowers whose colours were the only predictor of the nectar delivery. They examined how bees respond to the mean and variability of this delivery in a foraging version of a stochastic 'two-armed bandit' problem^{12,13}. In one series of experiments, all the blue flowers contained $2 \mu\text{l}$ of nectar, $\frac{1}{3}$ of the yellow flowers contained $6 \mu\text{l}$, and the remaining $\frac{2}{3}$ of the yellow flowers contained no nectar at all. In practice, 85% of the bees' visits were to the constant-yield blue flowers despite the equivalent mean return from the more variable yellow flowers. In a second series of experiments, Real showed that bumblebees will forage equally from each flower type if the mean reward from the variable type is made sufficiently large.

In the honeybee suboesophageal ganglion, an identified neuron, VUMmx1, delivers information about reward during classical conditioning experiments¹. This neuron projects widely to brain regions involved in odour processing, becomes active in response to sucrose applied to the antennae and proboscis, and its firing can substitute for the unconditioned stimulus in a classical conditioning experiment. Specifically, presentation of an

Received 11 May; accepted 23 August 1995.

- Suarez, R. K. *Experientia* **48**, 565–570 (1992).
- Wells, D. J. *J. exp. Biol.* **178**, 39–57 (1993).
- Hochachka, P. W. *Muscles as Molecular and Metabolic Machines* (CRC, Boca Raton, FL, 1994).

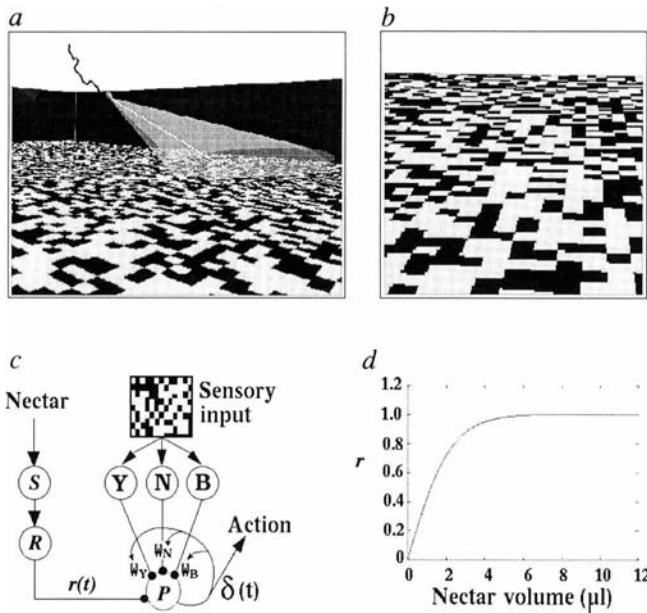


FIG. 1. Model of a foraging bee. *a*, Model bee moving about the three-dimensional arena along with a highlighted cone-shaped region indicating the field of view available to the simulated cyclopean eye. The arena was composed of blue and yellow squares (160 by 160) representing flowers of two different colours (shown here as black and grey). The model bee could move throughout the arena, though it reflected off the ceiling and walls. The trajectory of the bee is shown as a black trail. *b*, Image formed on the 200 by 200 pixel retina of the model bee illustrated in *a*. *c*, Architecture of the model. The output of neurons 'B', 'Y',

and 'N' represented changes in the percentage of blue, yellow and neutral colours in the visual field, hence their outputs reflect normalized responses. Such derivatives, typical of early sensory pathways, could be constructed anywhere along the path from the sensory units to *P*. Any portion of the field of view that did not contain blue or yellow was taken as neutral (white region in *b*). These neurons influenced the output of a diffusely projecting linear unit *P* through weights w_B , w_Y and w_N . w_B and w_Y were adaptable and w_N was fixed at -0.5 . As the model bee changed its heading and/or its height above the field, changes in the activity in these neurons would ensue. Changing the weights upon encounters with flowers permitted them to represent information about the predictive relationship between the sensory input and the amount of nectar^{19,23,30}. *d*, Response of neuron *R* that delivers information about the reward to *P* and receives input from a sucrose (nectar) sensitive neuron *S*. The functional form of this relationship is derived from an empirically determined utility function for bumblebees^{8,10}.

METHODS. The model bee had a single eye with a fixed field of view that was varied from 20° to 30° . This arrangement was not meant to represent the eye of the bee, but instead represents the use of visual information only from the central portion of the visual field along the direction of motion of the bee. At each iteration, the output of *P* influenced the decision of the bee to reorient randomly (Fig. 2). If no reorientation occurred, the bee took one step along its direction of flight, otherwise, it chose a random change in heading from -90° to $+90^\circ$ and then took a step (stepsize = 0.05). A landing was registered once the model bee's altitude was less than 0.05. At that point, the flower intersected by the direction of flight was selected. After landing on a flower, a reward is given according to the volume of nectar in the selected flower and activity in $r(t)$ resulted (*d*). The weights (w_B , w_Y) were adjusted only on encounters with flowers and otherwise influenced the decision of the bee to reorient through their influence on the fluctuating output of neuron *P*. After the model bee landed on a flower, received a nectar reward (or not), and had the sensory weights updated, it was randomly repositioned at the top of the simulated arena with a random initial heading. This explains the pattern of results shown in Fig. 4 at high altitudes.

odourant followed by artificial stimulation of VUMmx1 through an electrode caused subsequent presentation of the odourant alone to elicit a stereotypic behavioural response which indicated that the bee expects to receive nectar.

We have compared the behaviour of real bees to a neural model of bee learning based on VUMmx1 in a simulated three-dimensional arena containing a field of flowers possessing the same reward distributions as described above. In the model architecture, shown in Fig. 1, *P* was a simple linear unit receiving convergent sensory information representing the changes in the percentage of blue (x_B), yellow (x_Y) and neutral (x_N) inputs from the visual field, weighted by w_B , w_Y and w_N , respectively. In the presence of nectar (activity along $r(t)$), the output of the linear unit *P* was

$$\delta(t) = r(t) + \dot{V}(t) \equiv r(t) + V(t) - V(t-1) \quad (1)$$

where

$$V(t) = w(t) \cdot x(t) = w_B(t)x_B(t) + w_Y(t)x_Y(t) + w_N(t)x_N(t) \quad (2)$$

The output of *P*, $\delta(t)$, thus represents an ongoing comparison of $V(t-1)$ and the sum $r(t) + V(t)$, and, in the absence of reward, ($r(t) = 0$), *P*'s output labels transitions in sensory input as 'better than expected' ($\delta(t) > 0$) or 'worse than expected' ($\delta(t) < 0$). We interpret the sign of $\delta(t)$ as changes in neuromodulator release about some basal level^{14,15}.

After landing on a flower and receiving some volume of reward, the output of *P*, $\delta(t)$, controlled learning according to:

$$\Delta w(t) = \lambda x(t-1)\delta(t) \quad (3)$$

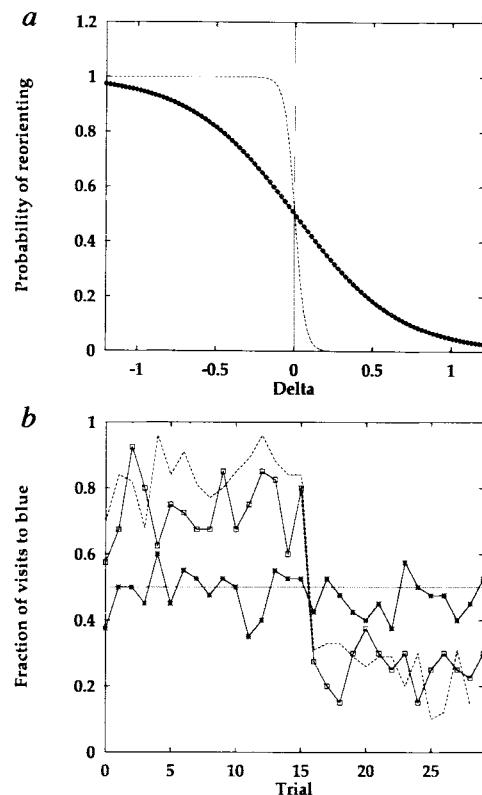
where λ is a learning rate. This learning rate is similar to one first proposed by Konorski¹⁶ and is a form of Hebb rule¹⁷ in which the postsynaptic factor $\delta(t)$ carries a sign permitting the system to learn predictions rather than correlations¹⁸⁻²². VUMmx1 is more highly activated by sensory cues that predict reward than by cues that do not predict reward and could have the same function as $\delta(t)$ in our model¹.

Once the model bee lands on a flower (Fig. 1), we use an empirically derived 'utility' curve^{8,10} to determine the value of different volumes of nectar (Fig. 1*c,d*). We interpret this curve as an equilibrium measure of the value of different volumes of nectar that does not specify detailed dynamics of how an individual bee might estimate nectar volumes or their subjective utility (see ref. 6). We made the simplifying assumptions that the utility of the nectar in a given flower could be assessed in one iteration and that the sensory component is $V(t) = 0$ as the model gathers nectar at time *t*. At this time $r(t)$ takes on the value prescribed by the utility curve and $V(t-1)$ is driven exclusively by the current flower colour. In this way, $\delta(t)$ takes on the form $r(t) + V(t) - V(t-1) = r(t) - V(t-1)$ making equation (3) equivalent to the Rescorla-Wagner rule for classical conditioning^{3-5,23}.

In the absence of reward ($r(t) = 0$), we used the output of *P* to bias actions. At each time *t*, $\delta(t)$ determined whether the bee continued on its present heading for its next movement or whether it randomly reoriented (tumbled) before its next movement (Fig. 2*a*). This model is reminiscent of a biased random walk with $\delta(t)$ choosing the probability of randomly changing direction. In chemotaxing bacteria, such decision-making is called klinokinesis²⁴. In various invertebrate systems, neuromodulator delivery influences motor behaviours and thus action choice²⁵⁻²⁷, however, we are making the novel proposal that VUMmx1 or its visually sensitive congeners also have such effects. The use of a signal such as $\delta(t)$ to learn to choose actions appropriately leads to an asynchronous version of dynamic programming²², an engineering technique that can be used to find optimal sequences of actions to achieve a goal.

In the results reported here, we assumed that weight changes only occurred during encounters with flowers. Thus, the output from *P* was used continuously to guide actions but plasticity was gated. This could occur by modulation of the learning rate (λ in equation (3)) by neurons that become active in response

FIG. 2. Simulations of bee foraging behaviour using predictive hebbian learning. *a*, Influence of $\delta(t)$ (labelled Delta) on the decision to reorient. The adaptable (w_B , w_V) and non-adaptable (w_N) sensory weights influence the decision to reorient through their influence on the size and sign of $\delta(t)$. $P(\delta(t))$ is the probability of reorienting for a given value of $\delta(t)$ and has the form $1/(1 + \exp(mx + b))$. The slope m of the linear region of this curve determines the amount 'noise' in the decision function: dashed line $m = 3$, line with points $m = 30$. In the results presented in this paper, m was varied from 5 to 45 and b varied from 0.1 to 5.0. *b*, Fraction of visits to blue flowers for real and model bees: weights updated only on flower encounters. Each trial represents about 40 flower visits averaged over 5 real bees and exactly 40 flower visits for a single model bee. Trials 1–15 for the real and model bees had blue flowers as the constant type ($2\mu\text{l}$ in all flowers), the remaining trials had yellow flowers as constant. Data from real bees are shown as dashed line. Data for simulated bees shown as connected points (learning rate $\lambda = 0.9$). The control trace (fluctuating trace around line at 0.5) shows the sampling behaviour for the model bee where $w_B = w_V = 0.5$ and no changes in the weights were made upon encounters with flowers. The control trace shows that the model for output is not biased toward either flower by other constraints associated with the chosen representation of the model bee and the arena. The real bees were more variable than the model bees and tended to sample the constant flowers at a slightly higher rate (85% compared to a range of 73%–85%). The real bees also had a slight preference for blue flowers which can be seen after the switch of reward distributions at trial 15 (ref. 7). For the model bee, the rewards were stochastically delivered so there was no effect of revisits on the effective variance of either the constant or variable flowers. In practice⁸, this assumption was not controlled, however, moderate increases in variance for the constant flower would not influence dramatically the behaviour of the model.



to signals specifically associated with a flower encounter, such as touch-sensitive neurons and odourant detectors with appropriate thresholds. Similar results were obtained when the learning rate was varied continuously according to the height of the bee above the field of flowers (unpublished data).

Figure 2*b* shows the behaviour of model bees compared with that of real bees⁸ in the experiment testing the extent to which they prefer a constant reward to a variable reward of the same long-term mean. The constant and variable flower types were switched after trial 15. The behaviour of the model matches best

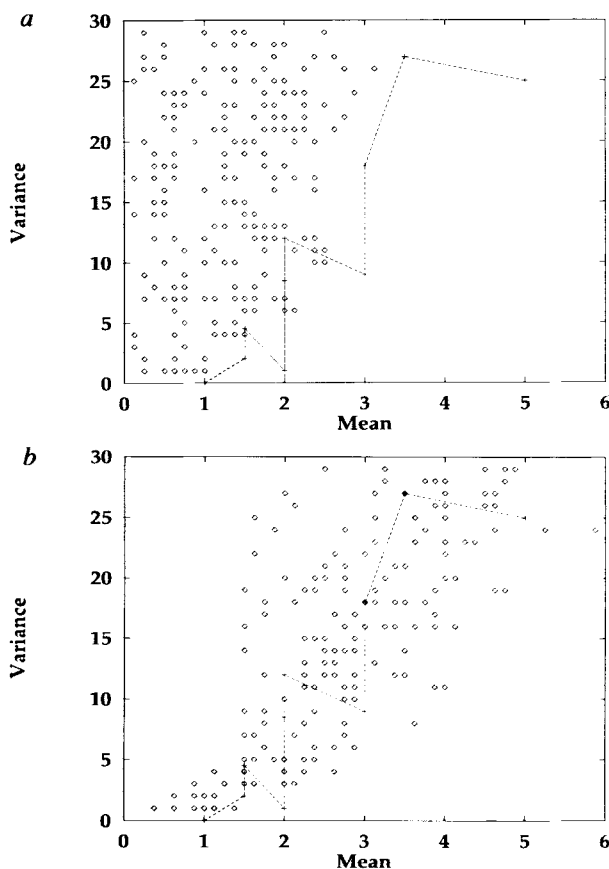


FIG. 3 Tradeoff between the mean and variance of nectar delivery for bee foraging. *a*, *b*, Compare indifference points for the model bees against those for real bees for different learning rates. The coordinates of indifference points are taken as the mean and variance for which the bee samples equally from both the constant and variable flowers. The constant flower contained $0.5\mu\text{l}$ of nectar. For the variable flower, the variance was fixed and the mean slowly increased until the percentage of visits to the constant flower was less than 50% over 80 flower visits: the mean and variance were then used as an indifference point. The stochastic nature of the model bee movement and the delivery of reward results in the observed spread in the indifference points. *a*, Indifference points for $\lambda = 0.05$. *b*, Indifference points for $\lambda = 0.9$. In both *a* and *b*, the data for real bees are shown as points connected by a solid line⁹ (units for mean and variance are μl and μl^2 , respectively).

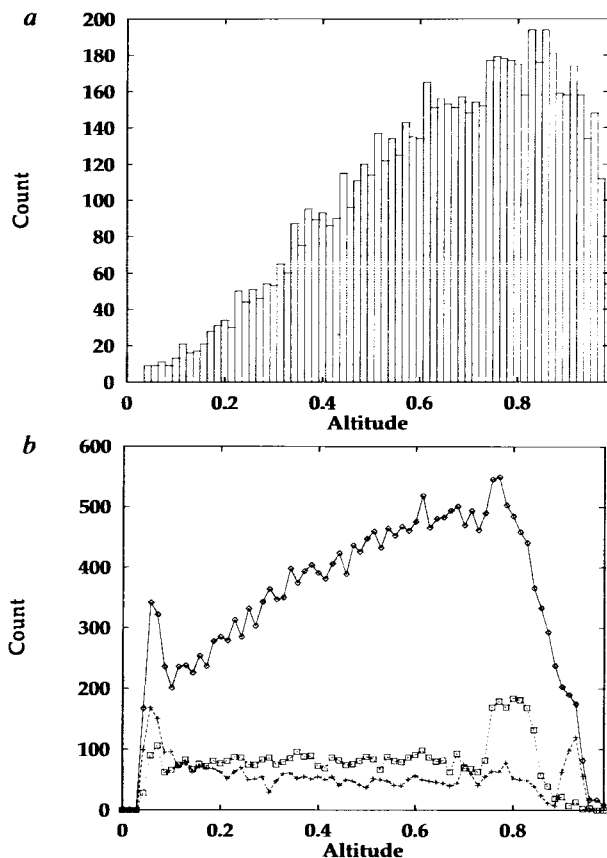


FIG. 4 Structure of foraging problem in an enclosed arena. The choice of a particular representation for the arena and the sensory apparatus of the model bee imposes a number of constraints that influence foraging in the simulated arena. The data presented in a and b are collected from a total of 1,200 flower landings and 41,735 steps of the model bee (stepsize=0.05). The learning rate λ was 0.9. a, Cumulative histogram of the number of times the model bee viewed primarily neutral colour ($>90\%$ of visual field) as a function of altitude in the arena. As the model bee gets closer to the field of flowers, this happens far less frequently. Since w_N was fixed at -0.5 , this histogram shows how often reorientations result from viewing regions outside the simulated field of flowers. b, Histograms of positive and negative fluctuations in $\delta(t)$ as a function of altitude. Top trace: distribution of all reorientations. Middle trace: distribution of negative fluctuations in $\delta(t)$ excluding those due to viewing primarily neutral colour ($\delta(t) < -0.1$). Bottom trace: distribution of positive fluctuations in ($\delta(t) > 0.1$). As the model bee approached the simulated field, a given change in orientation resulted in larger fluctuations in $\delta(t)$. In both panels, binsize=0.014 (70 bins).

the observed data for a learning rate of $\lambda=0.9$, suggesting that the real bee uses information over small time windows for controlling its foraging decisions⁸. These data show that for an equal sized reward between two behavioural choices, bees are biased to choose the more certain predictor of nectar. The uncertainty of a predictor can, however, be balanced by the expected size of the reward.

For example, as shown in Fig. 3, real bumblebees will forage equally on the constant and variable flower types if the mean reward from the variable type is made sufficiently large. This tradeoff behaviour was also exhibited by the model bee. For a given variance in reward from the variable flower, the mean reward was increased until the model bees foraged equally from each flower type: at that point the mean and variance was recorded as the point at which the model bee sampled indifferently from each type. Figure 3a,b show such indifference plots for two learning rates, 0.05 (a) and 0.9 (b). As before, the behaviour of real bees was matched best by the higher learning rate (0.9). The reason for the shape of the observed plots can be traced to the nonlinear response function in Fig. 1d, which provides diminishing returns for larger volumes of nectar.

This tradeoff between the expected value of reward and its uncertainty is not, however, universal. In choice experiments in honeybees, Greggers and Menzel⁶ showed that honeybees can be induced to maximize their return by choosing a single flower. This paper was one of the first to show how learning might affect foraging based on the Rescorla–Wagner model of conditioning. We extend this view, showing how the learned weights can be used to choose appropriate actions and how the resulting action choices influence the learning. In addition, the actions taken by the model are also influenced by the structure of the simulated environment in which it moved. Figure 4 shows how the output of the diffuse neuron P was influenced by the structure of the foraging problem faced by the model bee. Hence, although learning (weight changes) was one major influence in the behaviour exhibited by the model, the structure of the environment played an important role in shaping its behavioural decisions.

The success of the model in accounting for the choice behaviour of bumblebees allows us to connect the performance of a simple neural system using known anatomical and physiological constraints and descriptions of a behaviour previously explained mainly in terms of a decision-theoretical framework for minimizing risk (ref. 8 but see ref. 6). Visual, gustatory and VUMmx1 inputs converge on the antennal lobes and the mushroom bodies, thus making these regions good candidate sites for the learning rule in equation (3). Delivery of octopamine, which is released by VUMmx1, directly to these regions can substitute for the firing of VUMmx1 in some conditioning experiments²⁸.

There is good evidence for similar predictive responses in primate mesencephalic dopaminergic systems^{14,15,29}. Hence, dopamine delivery in primates may be used by target neurons to guide action selection and learning, suggesting the conservation of an important functional principle, albeit differing in its detailed implementation. □

Received 27 February; accepted 18 August 1995.

1. Hammer, M. *Nature* **366**, 59–63 (1993).
2. Hammer, M. & Menzel, R. *J. Neuroscience* **15**, 1617–1630 (1995).
3. Couvillon, P. A. & Bitterman, M. E. *Anim. Learn. Behav.* **13**, 246–252 (1985).
4. Couvillon, P. A. & Bitterman, M. E. *Anim. Learn. Behav.* **14**, 225–231 (1986).
5. Couvillon, P. A., Lee, Y. & Bitterman, M. E. *Anim. Learn. Behav.* **19**, 381–387 (1991).
6. Greggers, U. & Menzel, R. *Behav. Ecol. Sociobiol.* **32**, 17–29 (1993).
7. Menzel, R., Greggers, U. & Hammer, M. in *Insect Learning: Ecology and Evolutionary Perspectives* (eds Papaj, D. R. & Lewis, A. C.) 79–125 (Chapman and Hall, London, 1993).
8. Real, L. A. *Science* **253**, 980–986 (1991).
9. Real, L. A. *Ecology* **62**, 20–26 (1981).
10. Harder, L. D. & Real, L. A. *Ecology* **68**, 1104–1108 (1987).
11. Real, L. A., Ellner, S. & Harder, L. D. *Ecology* **71**, 1625–1628 (1990).
12. Berry, D. A. & Fristedt, B. *Bandit Problems: Sequential Allocation of Experiments* (Chapman and Hall, London, 1985).
13. Krebs, J. R., Kacelnik, A. & Taylor, P. *Nature* **275**, 27–31 (1978).
14. Montague, P. R., Dayan, P., Nowlan, S. J., Pouget, A. & Sejnowski, T. J. in *Advances in Neural Information Processing Systems 5* (eds Hanson, S. J., Cowan, J. D. & Giles, C. L.) 969–976 (Morgan Kaufmann, San Mateo, CA, 1993).
15. Montague, P. R. & Sejnowski, T. J. *Learn. Memory* **1**, 1–33 (1994).
16. Konorski, J. *Conditioned Reflexes and Neuron Organization* (Cambridge University Press, Cambridge, 1948).
17. Hebb, D. O. *The Organization of Behavior* (Wiley, New York, 1949).
18. Samuel, A. L. *IBM. J. Res. Dev.* **3**, 211–229 (1959).
19. Sutton, R. S. & Barto, A. G. *Psych. Rev.* **88**, 135–170 (1981).
20. Sutton, R. S. & Barto, A. G. in *Learning and Computational Neuroscience* (eds Gabriel, M. & Moore, J. W.) 497–538 (MIT Press, Cambridge, MA, 1987).
21. Sutton, R. S. *Machine Learn.* **3**, 9–44 (1988).
22. Barto, A. G., Sutton, R. S. & Watkins, C. J. C. H. *Technical Report 89-95*, (Computer and Information Science, University of Massachusetts, Amherst, MA, 1989).
23. Rescorla, R. A. & Wagner, A. R. in *Classical Conditioning II: Current Research and Theory* (eds Black, A. H. & Prokasy, W. H.) 64–69 (Appleton-Century-Crofts, New York, 1972).
24. Spudis, J. L. & Koshland, D. E. *Proc. natn. Acad. Sci. U.S.A.* **72**, 710–713 (1975).
25. Lockery, S. R. & Kristan, W. B. *J. comp. Physiol.* **168**, 165–177 (1991).
26. Goldstein, R. S. & Camhi, J. M. *J. comp. Physiol.* **168**, 103–112 (1991).
27. Harris-Warrick, R. M., Coniglio, L. M., Barazangi, N., Guckenheimer, J. & Gueron, S. J. *Neurosci.* **15**, 342–358 (1995).
28. Hammer, M. & Menzel, R. *Soc. Neurosci. Abstr.* **20**, 582 (1994).
29. Schultz, W., Apicella, P. & Ljungberg, T. *J. Neurosci.* **13**, 900–913 (1993).
30. Widrow, B. & Stearns, S. D. *Adaptive Signal Processing* (Prentice-Hall, Englewood Cliffs, 1985).

ACKNOWLEDGEMENTS. We thank P. S. Churchland, A. Dayan, A. Pouget, D. Raizen, S. Quartz and R. Zemel for their helpful comments and criticisms on earlier versions of this work. We also thank P. Yates and D. Egelman for help with computer simulations. This work was supported by the National Institute of Mental Health (P.R.M., T.J.S.), Center for Theoretical Neuroscience at Baylor College of Medicine (P.R.M.), Howard Hughes Medical Institute (T.J.S.), Natural Science and Engineering Research Council (Can) (P.D.), Science and Engineering Research Council (UK) (P.D.).

Seeing motion behind occluders

Scott N. J. Watamaniuk* & Suzanne P. McKee

The Smith-Kettlewell Eye Research Institute, 2232 Webster Street, San Francisco, California 94115, USA

THE visual system has no difficulty maintaining the identity of an object as it disappears and reappears behind stationary occluders. In the natural world, a moving object may differ from occluders by many characteristics (colour, depth, shape and so on). Scene segmentation based on these characteristics is thought to happen early in visual processing, and to influence how objects, including moving objects, are identified¹⁻⁵. What happens if the only characteristic distinguishing an object is its direction of motion? Experiments with random dot displays show that one dot moving in a constant trajectory is readily detected among identical dots in brownian motion⁶. Detection declines sharply if the trajectory is intermittently broken, but improves if occluders obscure the breaks in the trajectory. It is not sufficient that these occluders be perceived as segmented from the rest of the display (such as by colour or depth). Rather, it is critical that the occluders do not contain motion that is similar in direction to that of the target trajectory. We conclude that detection of the trajectory is due to the integration of information within a network of low-level motion detectors and is not dependent on segmentation processes⁷.

We investigated how the detectability of one dot moving in a constant direction embedded in a background of identical dots in brownian motion would be affected by the presence of various types of occluders. Stimuli were computer-generated random-dot cinematograms viewed through a circular aperture (9 deg in diameter). The display was implicitly divided equally into nine vertical strips and the motion of dots in alternate strips was controlled independently; dots appearing in the even strips could move according to one algorithm, while dots appearing in the odd strips could move according to a different algorithm. In our experiments, the odd strips always contained dots in brownian motion while the characteristics of the even strips were varied. In the central three odd strips, amidst the brownian motion, one 'signal dot' moved in a constant direction. The motion of this signal dot was such that after leaving one odd strip, and being extinguished, it reappeared in the next odd strip precisely at a position and time as if it had moved through the intervening strip on its previous trajectory (Fig. 1a). The direction and position of the trajectory could vary from trial to trial (Fig. 1b).

We created seven displays in which the characteristics of the even strips, where the signal trajectory did not appear, differed. The odd vertical strips in which the 'signal trajectory' was presented always contained dots moving in brownian motion. Detectability of the signal trajectory under the seven experimental conditions is shown in Fig. 2. When the even strips contained dots moving in random directions (top left), detection was poor, producing an average $d' = 0.47$ (63% correct detection). However, when the even strips were covered with an opaque mask or contained dots moving upwards, detectability of the signal trajectory improved to $d' = 1.14$ (79% correct detection). How can this improvement in detectability be explained?

The temporal coherence model^{7,8} uses interconnections between low-level motion detectors to account for psychophysical data on the detectability of trajectory motion embedded in brownian motion noise⁹. When a motion detector is stimulated, the model averages (smooths) this response with the responses of other spatially adjacent detectors with similar directional tuning. If only one detector in the spatial averaging region has a large signal, its signal will be reduced by the averaging process. However, temporal averaging also occurs. If a detector at time t is stimulated and at time $t + 1$ an adjacent detector that lies along the tuned direction is stimulated, the adjacent detector

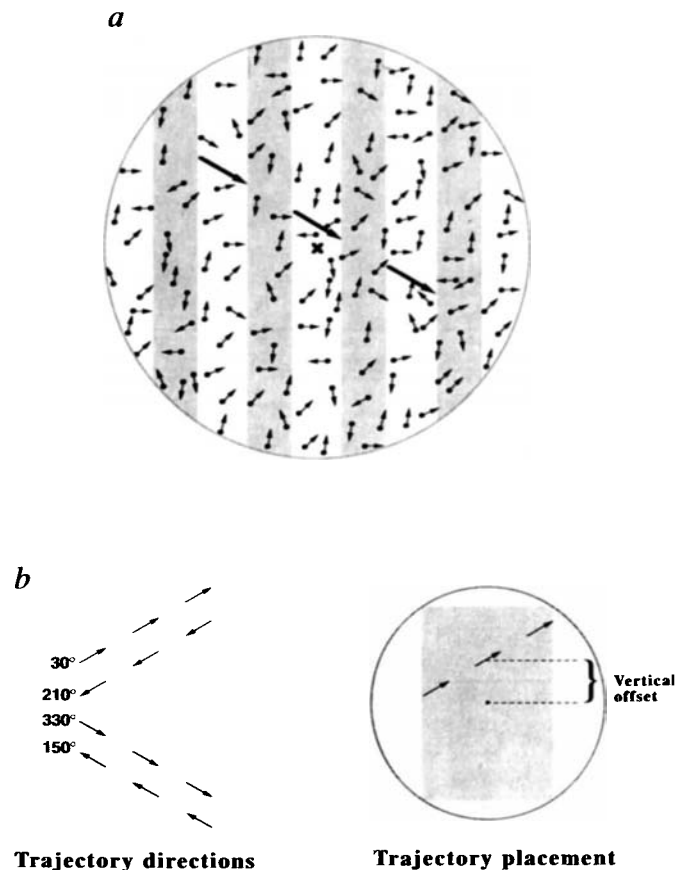


FIG. 1 a, A schematic representation of our random-dot stimulus. Shading is used here to illustrate the virtual strips of the display but the strips were not demarcated in any way in the experimental stimuli unless otherwise indicated. Stimuli were presented at a rate of 50 Hz on an x-y cathode-ray-tube display. Dot density was constant at 1.5 dots deg⁻². Noise dots were not allowed to move from one strip to another. If a noise dot's motion forced it outside its original strip, the dot was randomly positioned back inside the strip. Stimulus duration was 600 ms for all conditions. The time course of the signal dot was as follows. The signal dot appeared at the far end of an odd strip 50 ms after stimulus onset and completely traversed the strip in its assigned direction in 100 ms. The signal dot then became invisible for 100 ms and reappeared at the nearest edge of the next odd strip precisely at the location as if it had moved through the intervening even strip at its original speed and direction. This was repeated until the signal dot moved through the three most central odd strips. Stimulus offset followed 50 ms after the signal dot exited the third odd strip. b, The signal dot moved in one of four possible oblique directions, each 30 deg off horizontal, and was centred horizontally on the display but with a vertical offset between ± 2 deg randomly determined each trial. A centred stationary fixation spot was always visible. All moving dots were displaced with the same frame-to-frame spatial step as the 'signal dot' (0.2 deg) and so all dots moved at the same speed (10 deg s⁻¹) in their respective directions.

will have a larger signal than if the motion at time t had not occurred. Because signals will steadily increase in those detectors responding to consistent-direction motion, this model has been referred to as a 'trajectory network'. Thus, according to the model, our displays produce two countervailing effects. The consistent direction of the signal dot enhances detectability through temporal averaging, but the surrounding noise dots degrade detectability through spatial averaging, because adjacent motion detectors, with directional tuning similar to the detector responding to the signal dot, receive weak stimulation from noise dots.

When small trajectory segments are separated by brownian motion noise, as in our stimuli, the process of temporal averaging is disrupted. Moreover, the spatial averaging process will

* Present address: Psychology Department, Wright State University, Dayton, Ohio 45435, USA.

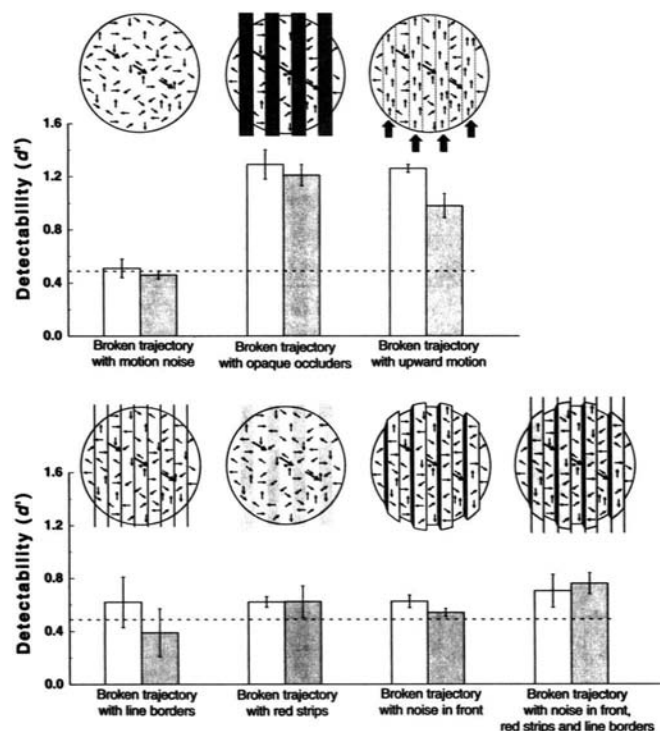


FIG. 2 Detection of the trajectory motion target for seven experimental conditions for two observers, represented by the grey and white bars. Experimental conditions differ by the characteristics of the even strips. The even strips contained either dots moving in brownian motion, an opaque occluding mask (a piece of black cardboard placed in front of the display), dots moving upwards, dots moving in brownian motion but with solid lines demarcating strip boundaries, dots moving in brownian motion but covered with a piece of red transparency, dots moving in brownian motion but presented at a depth 13 arcmin in front of the odd strips, which were in the fixation plane, or dots moving in brownian motion covered with a piece of red transparency presented at a depth 13 arcmin in front of the odd strips with black lines demarcating strip boundaries. These stereoscopic displays were created by presenting a separate image to each eye by using a pellicle and polarizing filters. In all stimuli, the odd vertical strips in which the 'signal trajectory' was presented had dots moving in brownian motion. A schematic diagram above each pair of columns represents the stimulus condition. Each data point is the result of at least 100 trials. Error bars represent ± 1 s.e.m. Detectability of the signal dot was measured using a temporal two-interval forced-choice procedure. In each trial, the observer was shown two stimuli (600 ms each) of which only one contained the signal trajectory. The task of the observer was to identify which stimulus interval contained the signal trajectory. The broken line in each graph is provided as a reference, and is the average performance in the 'motion noise' condition (top left).

still degrade the signal in the trajectory network because the noise strip contains dots moving in directions similar to that of the trajectory, and some of these will appear in the vicinity of the implicit path of the trajectory. Because the signals in the detectors responding to the brownian motion are small, averaging these small signals with the trajectory's larger signal will reduce the trajectory's signal. Under these conditions, the trajectory network must start afresh for each new trajectory segment, resulting in poor detectability. This explains the low detectability of the three-segment signal trajectory when the even strips contain brownian motion (first pair of data columns in Fig. 2).

The trajectory network can also account for our results when the even strips were occluded or contained upward motion (see Fig. 2). When the even strips were masked by opaque occluders, detectability improved. Presumably, when no motion detectors are stimulated, as in the occluded regions, the large signals within the trajectory network can continue to propagate for a limited amount of time and space. When this propagated signal reaches the next visible strip and the trajectory motion continues, the

signals of the detectors responding to the trajectory are again enhanced by the signals generated by the previous trajectory segment. Thus signals in detectors responding to the trajectory motion continue to grow and detectability improves. The same process occurs when the even strips contain upward motion. Because spatial averaging within the trajectory network occurs only between adjacent detectors responding to similar directions of motion, the dissimilar upward motion does not influence the signal resulting from the trajectory's motion.

A plausible alternative explanation of our results is that whenever the even strips can be segregated from the odd strips, the even strips are interpreted by the visual system as occluders. Shipley and Kellman¹⁰ proposed that perturbations of local elements are classified, using a continuity rule, to determine if their motions are continuous or discontinuous with nearby elements. Discontinuity or optic tearing indicates a surface boundary. The continuity constraint embodied in the trajectory network may be the same as that proposed by Shipley and Kellman.

We tested the 'segregation hypothesis' with four additional displays in which the even strips contained brownian motion but with various manipulations to make the even strips distinct from the odd strips. One display had visible black lines at the borders between strips, one had a red transparency placed over the even strips, one had the even strips presented stereoscopically at a disparity 13 arcmin in front of the odd strips, and one combined all three characteristics. The data in Fig. 2 (bottom) show that none of these four conditions produced the enhanced detectability as in the opaque occluder or upward-motion conditions. This is surprising because previous research had shown depth to influence the perceived direction of motion in competing-motion displays^{3,4}. In our experiments, these displays were clearly perceived to have strips in two separate depth planes and yet detectability of the trajectory in the odd strips was similar to the initial broken-trajectory condition. These results suggest that, even though motion may be perceived to belong to separate objects or regions even at different depths, the mechanism detecting trajectory motion⁹ does not reflect this organization.

Motion in the noise strips that is similar in direction to the trajectory disturbs or adds noise to the trajectory network. It is the removal of this specific-motion noise that results in higher rates of detection. Hence, when the even strips contained coherent upward motion, sufficiently dissimilar from the trajectory directions, detection of the trajectory was almost as good as when the even strips were occluded. In support of this view, we found clear evidence of directional tuning under the present experimental conditions. When the even strips contained motion in the same direction as one of the target trajectories, detectability of that trajectory direction was reduced to very near chance ($d' = 0.1$, 53% correct detection) but the other three trajectory directions showed enhanced detectability ($d' = 1.05$, 77% correct detection).

Our data are in accordance with the existence of trajectory networks composed of interconnected motion detectors that utilize a continuity constraint (direction-selective averaging) to enhance the signal produced by trajectory motion^{6,9,11}. The continuity constraint in the trajectory network and that underlying segmentation¹⁰ are different. □

Received 26 June; accepted 30 August 1995.

1. Nakayama, K. & Shimojo, S. *Vision Res.* **30**, 1811–1825 (1990).
2. Shimojo, S., Silverman, G. H. & Nakayama, K. *Nature* **333**, 265–268 (1988).
3. He, Z. L. & Nakayama, K. *Nature* **367**, 173–175 (1994).
4. He, Z. L. & Nakayama, K. *Nature* **369**, 231–233 (1992).
5. Braddick, O. J. *Nature* **333**, 803–804 (1988).
6. Watamaniuk, S. N. J., McKee, S. P. & Grzywacz, N. M. *Vision Res.* **35**, 65–77 (1995).
7. Yuille, A. L. & Grzywacz, N. M. *Nature* **333**, 71–74 (1988).
8. Grzywacz, N. M., Smith, J. A. & Yuille, A. L. in *The Proceedings of the IEEE Workshop on Visual Motion 148–155* (IEEE Computer Society, 1989).
9. Grzywacz, N. M., Watamaniuk, S. N. J. & McKee, S. P. *Vision Res.* **325** (in the press).
10. Shipley, T. F. & Kellman, P. J. *Spatial Vis.* **7**, 323–339 (1993).
11. Bowne, S. F., McKee, S. P. & Glaser, D. A. *J. opt. Soc. Am. A* **6**, 1112–1121 (1989).

ACKNOWLEDGEMENTS. We thank O. Braddick for helpful comments on an early version of this manuscript. This work was supported by the U.S. Air Force Office of Scientific Research.

Responses of cells in monkey visual cortex during perceptual filling-in of an artificial scotoma

Peter De Weerd*, Ricardo Gattass†, Robert Desimone‡ & Leslie G. Ungerleider*§

* Laboratory of Psychology and Psychopathology and ‡ Laboratory of Neuropsychology, NIMH, Bethesda, Maryland 20892, USA

† Department of Neurobiology, Biophysics Institute Carlos Chagas, Federal University of Rio de Janeiro, Rio de Janeiro, RJ 21941, Brazil

WHEN we view a scene through one eye, we typically do not see the scotomas created by the optic disc and the blood vessels overlying the retinal surface. Similarly, when a texture field containing a hole is steadily viewed in peripheral vision (artificial scotoma), the hole appears to fill in with the surrounding texture in a matter of seconds, demonstrating that the visual system fills in information across regions where no information is available¹⁻⁵. Here we show that, in monkeys viewing a similar texture field with a hole, the responses of extrastriate visual neurons with receptive fields covering the hole increase gradually to a level comparable to that elicited by the same texture without a hole. The time course of these dynamic changes in activity parallels the time course of perceived filling-in of the hole by human observers, suggesting that this process mediates perceptual filling-in.

The receptive fields of visual area V1 neurons in anaesthetized cats have been shown to expand after several minutes of stimulation with a dynamic texture surrounding the receptive field, suggesting that expansion of the receptive field might underlie filling-in⁶. However, psychophysical studies in humans have shown that perceptual filling-in of a hole in a texture (artificial scotoma) occurs within seconds rather than minutes³⁻⁵. We re-investigated the neural basis of perceptual filling-in by comparing the time course of filling-in in human subjects with the time course of neuronal responses recorded from monkeys, both observing the same stimuli.

First, we determined the time course of perceptual filling-in in four human subjects. We used a large ($16^\circ \times 16^\circ$) texture with an equiluminant hole in its middle, located 8° from a fixation spot (Fig. 1). While maintaining fixation, subjects indicated when they saw the hole fill in. As the hole size was increased from 1.0° to 12.8° , the time required to see the hole fill in increased steadily. We next recorded from neurons in two rhesus monkeys that were rewarded for maintaining fixation, while the same stimuli as were used with human subjects were presented. For each cell, we centred the hole over the receptive field (average eccentricity of 8°). Responses to the texture with a hole (hole condition) were compared with responses to the same texture without a hole (no-hole condition). The monkeys were not required to signal the filling-in of the hole, but there is behavioural evidence for filling-in across the blind spot in this species⁷.

Physiological recordings in visual areas V2 and V3 indicated cells whose firing rate in the hole condition was initially lower than in the no-hole condition but gradually increased to a comparable level; that is, after a few seconds of fixation they responded to the texture with the hole as if it were a texture without a hole. An example of such a cell in V3 is shown in Fig. 2a; it responded with a sustained high firing rate to a continuous texture. By contrast, when the receptive field was covered by a hole of 4° in the texture, the cell gave a transient response to stimulus onset followed by a low firing rate, which then 'climbed' to a level comparable to that in the no-hole condition. We term

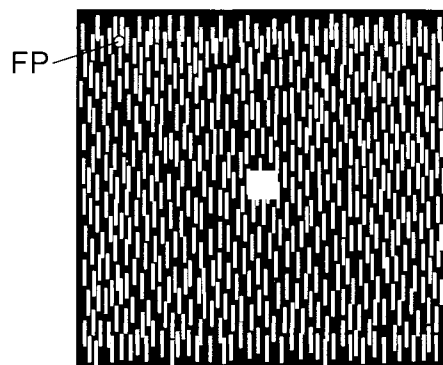


FIG. 1 Schematic representation of the stimulus, which consisted of three frames played at 20 Hz for 12 s. Each frame consisted of small bars ($0.2^\circ \times 1.4^\circ$), randomized in position, and spaced 0.9° apart on average. The hole was equiluminant to both the texture and the background (23 cd m^{-2}). From a short distance, the reader might experience filling-in of the small hole while steadily fixating the fixation point (FP), even though, unlike the experimental stimulus, the hole is more luminant than the texture and the texture is static, conditions which delay filling-in³⁻⁵.

METHODS. Human observers fixated the fixation point at the centre of the display and signalled when they saw the hole fill in by releasing a button. Trials per subject for each hole size ranged from 10 to 40. In both monkeys, areas V2 and V3 were localized by previous physiological mapping^{19,20}, and in one monkey the recording sites were confirmed with magnetic resonance imaging (MRI) scans taken with tungsten electrodes (clearly visible in the scans) located in typical recording sites. Standard physiological techniques were used, and eye movements were monitored with the scleral coil technique²¹. The monkeys were rewarded with fruit juice for maintaining fixation. To allow for occasional movements caused by eye blinks, the fixation window was set at 2.5° for one monkey and 3.5° for the other. Trials with eye movements outside those ranges were aborted and discarded during data analysis. The standard deviation of gaze from the fixation point was 0.39° in one monkey and 0.79° in the other, with no change in average eye position during the trial. The eye was located within 1° of the fixation point for 97% of the total trial time in one monkey and 92% of the time in the other.

this phenomenon 'climbing activity'. The climbing activity occurred within a time span that was comparable to the time course of perceptual filling-in; that is, the time required by human subjects to perceive the hole fill in (range 6.2–9.8 s) corresponded well with the time period during which the climbing activity approached the level in the no-hole condition (Fig. 2a). Significant climbing activity was observed for 10 of 35 cells in V2, and 68 of 127 cells in V3 (see Fig. 3a legend). Similarly, the distribution of response changes during filling-in across the entire population showed a shift towards response increases, which was larger in V3 than in V2 (Fig. 2b). These results indicate that the perception of filling-in resulted from a minimization of response differences in the hole and no-hole conditions.

The time during which the neuronal responses in the hole and no-hole conditions converged matched the time course of perceived filling-in at other hole sizes as well. The average response of cells with climbing activity in areas V2 and V3 is shown in Fig. 3a, for 93 cells tested with the full range of hole sizes. As the hole size increased to 5.6° , the initial transient response to stimulus onset decreased, subsequent activity was reduced, and more time was required for the climbing activity to reach the activity level in the no-hole condition. For hole sizes up to 5.6° , the time at which neuronal responses in V2 and V3 reached a minimal difference between hole and no-hole conditions fell within the range of times at which human subjects reported filling-in (Fig. 3b). For a hole size of 12.8° , partial filling-in limited to the hole's corners was occasionally reported. Similarly, climbing activity was strongly reduced and did not reach the firing rate in the no-hole condition (Fig. 3a).

§ To whom correspondence should be addressed.

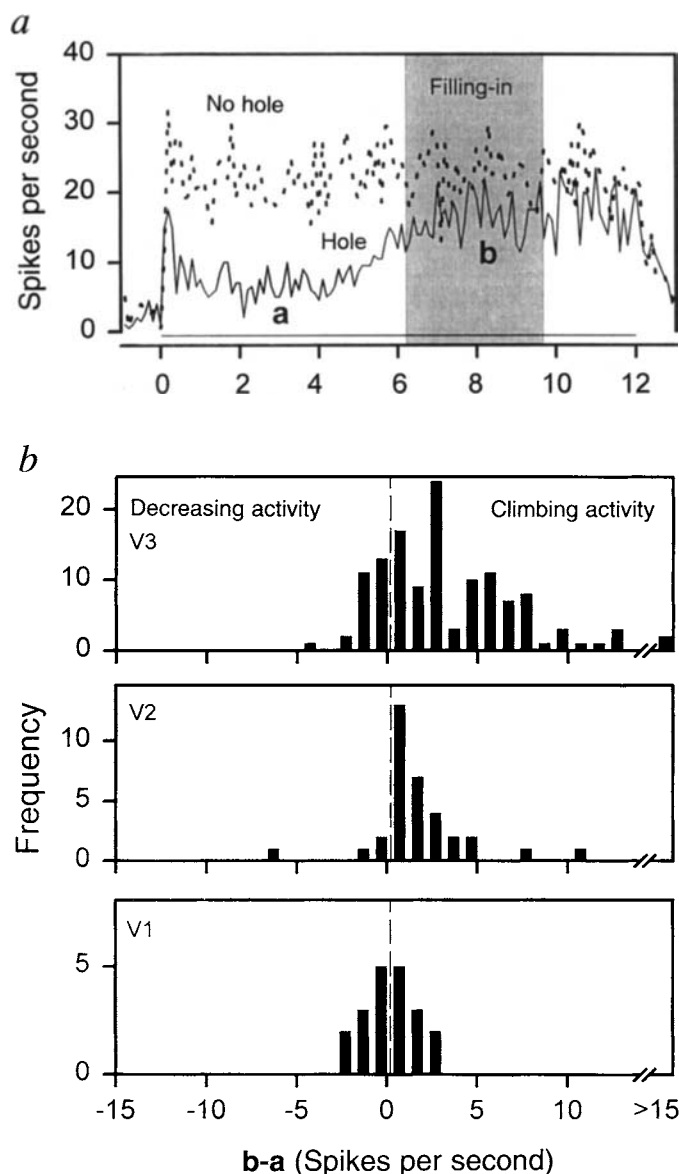


FIG. 2 Responses of cortical neurons during perceptual filling-in. **a**, Typical responses (binwidth, 100 ms) in the hole condition (solid line) and no-hole condition (dotted line) from a single V3 cell. Conditions were randomly interleaved, with up to 20 trials in each. The horizontal line indicates the 12-s stimulus presentation time, preceded and followed by a 1-s period in which activity was recorded in the absence of a stimulus. The hole in the texture was 4° , centred over the receptive field (measured by minimum response field method) which was 3.5° in size at an eccentricity of 7.2° . The orientation of the bars and square hole were typically chosen to match the preferred orientation of the cell. Preliminary data indicate that climbing activity also occurs for orthogonal orientations. The shaded zone indicates the range of average times required by human observers to report filling-in. **b**, Distribution of response changes in V3, V2 and V1 during perceptual filling-in (4° hole). Response changes were calculated for each cell by subtracting the firing rate in a period extending from 1.5 to 4.0 s after stimulus onset (labelled **a** in Fig. 2a) from the firing rate during perceptual filling-in (labelled **b** in Fig. 2a). According to a Wilcoxon paired-comparison test, there was a significant increase in activity (**b-a**) across the population in V3 ($P < 0.001$) and in V2 ($P < 0.001$), but not in V1 ($P = 0.82$). At the time filling-in occurred (shaded zone), there was no significant difference between the activity in hole and no-hole conditions in the V3 population ($P = 0.31$) although a significant difference remained in V2 ($P < 0.001$) and in V1 ($P < 0.001$). Data shown for V1 are from cells with an average eccentricity of 24.5° , where the human subjects experienced strong filling-in. The average receptive field size (2.5°) at this eccentricity was comparable to that of cells in V2 (2°) and V3 (3.7°) at the smaller eccentricities (9.1° and 7.6° , respectively) studied in these areas. Results from 13 V1 cells recorded at an eccentricity of 4.5° (not shown) were similar to those obtained at a 24.5° eccentricity. The distribution and statistical tests in V2 do not include one cell which exhibited inhibitory responses, its behaviour being the mirror image of the cells with climbing activity.

Other correlations between the behaviour of the human observers and the behaviour of the cells were then explored. With a hole of 4° , perceptual filling-in could be prevented by shrinking the overall size of the texture to a $4.4^\circ \times 4.4^\circ$ field, that is, by creating a narrow frame of texture around the hole. We tested this stimulus on seven V3 cells that showed climbing activity with the standard large texture, but none of them exhibited climbing activity with the reduced field (Fig. 4a). Therefore, the hole's boundaries *per se* do not cause either climbing activity or filling-in; rather, both require a sufficiently large field of texture in the receptive field surround. We could also prevent perceptual filling-in and climbing activity by flickering the texture field surrounding the hole at a rate of 1.9 Hz (Fig. 4b).

Limited data in V1 indicated only weak climbing activity. Significant climbing activity was observed for only 4 of 20 cells recorded at an eccentricity of 24.5° , and 0 of 13 cells recorded at an eccentricity of 4.5° . Furthermore, there was no overall increase in response across the V1 population during filling-in at either eccentricity, unlike the case in V2 and V3 (Fig. 2b). The data show that climbing activity is more pronounced and more prevalent in extrastriate cortex than in V1, and within extrastriate cortex it appears to be stronger and more prevalent in V3 than in V2 (see Fig. 2b).

To test whether climbing activity is due to expansion of the receptive field⁶, we quantitatively mapped the receptive field and

surround with moving or flashing light bars presented before, during, and after stimulation with the texture display. In some experiments we used two bar contrasts (Michelson indices of 23% and 67%) to test for changes in contrast gain⁸. In the 16 V3 cells with climbing activity that were tested, there were no changes in the size or profile of the receptive field, which indicates that increases in neither receptive field size nor contrast gain mediate climbing activity or filling-in (Fig. 4c, d). We suggest that climbing activity is caused by a dynamic change in the balance between excitation and inhibition in the receptive field and surround^{9,13}. The excitatory onset responses to the texture with the hole, even with large hole sizes, indicate that the cells receive excitatory input from surrounding regions extending well beyond the classical receptive field. The subsequent sharp fall-off of activity suggests that these excitatory inputs are followed by strong inhibition (Fig. 3d). We propose that the inhibition adapts over time, unmasking previously ineffective excitation from the texture, causing climbing activity. The closer the texture is to the centre of the receptive field, the stronger are the excitatory inputs, which explains why climbing activity occurs faster with smaller hole sizes. When the neuronal response to the surround texture becomes comparable to the response to a continuous texture, the subject experiences filling-in. Both long-range horizontal connections between pyramidal cells¹⁴⁻¹⁶ and feedback connections¹⁷ between areas may underlie the interaction

FIG. 3 Correlation between the time course of climbing activity and that of perceptual filling-in. *a*, Average activity in the hole condition (solid line) and no-hole condition (heavy dotted line) of V2 and V3 neurons with significant climbing activity. Fine dots show baseline activity recorded without any stimulus. Hole size is given at the top of the panels. Range of average times required by observers to report filling-in is indicated by the shaded zones. Other conventions as in Fig. 2. Climbing activity was evaluated with paired *t*-tests (two-tailed, $P < 0.05$) in which firing rates in a 2.5-s interval starting 1.5 s after stimulus onset were compared with the rates in the last 2.5 s of stimulus presentation. A few cells that showed a response increase in the no-hole condition were excluded from the count of cells with climbing activity. For cells that initially responded similarly in the hole and no-hole conditions there was less opportunity to develop climbing activity. This was particularly true at hole sizes of 3.2° and smaller (which were not as conspicuous to the subjects as the larger holes), for which we found significant climbing activity in only 12% of the cells. *b*, Comparison between the average time required by humans to report filling-in (open symbols) and the average time it took neurons with climbing activity to minimize the response difference in the hole and no-hole conditions (solid symbols). Data from individual observers are shown with thin, solid lines. The slopes of linear regression lines fit through the observers' data ranged from 0.8 to 1.1 ($r^2 > 0.90$ for three subjects and $r^2 = 0.73$ for the fourth) compared with 0.9 for the cells ($r^2 = 0.94$). For the neuronal data, the smallest difference between the firing rates in the hole and no-hole conditions was determined for each 1-s period beginning 1 s after stimulus onset, for each cell and hole size. The time at which the difference was smallest was considered as the 'filling-in time' for that cell at that hole size. Data at the smallest hole size were pooled from hole sizes of 1.0° and 1.6° .

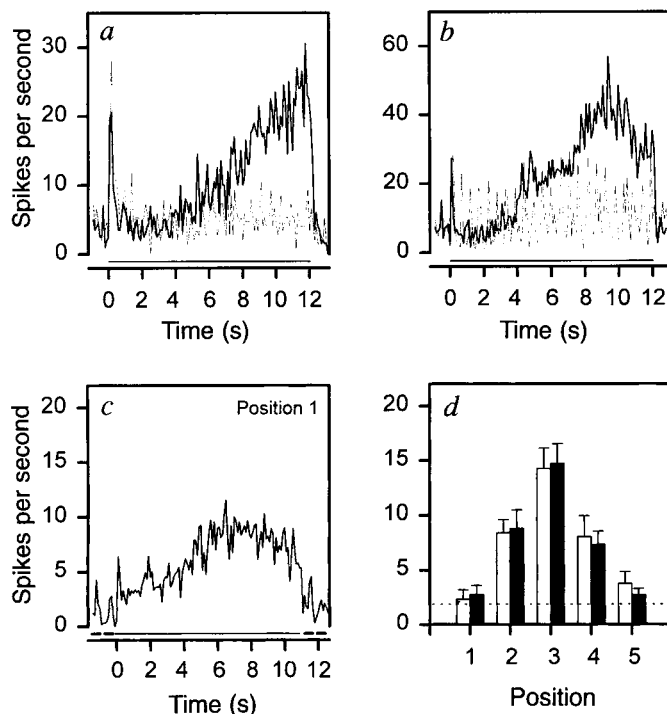
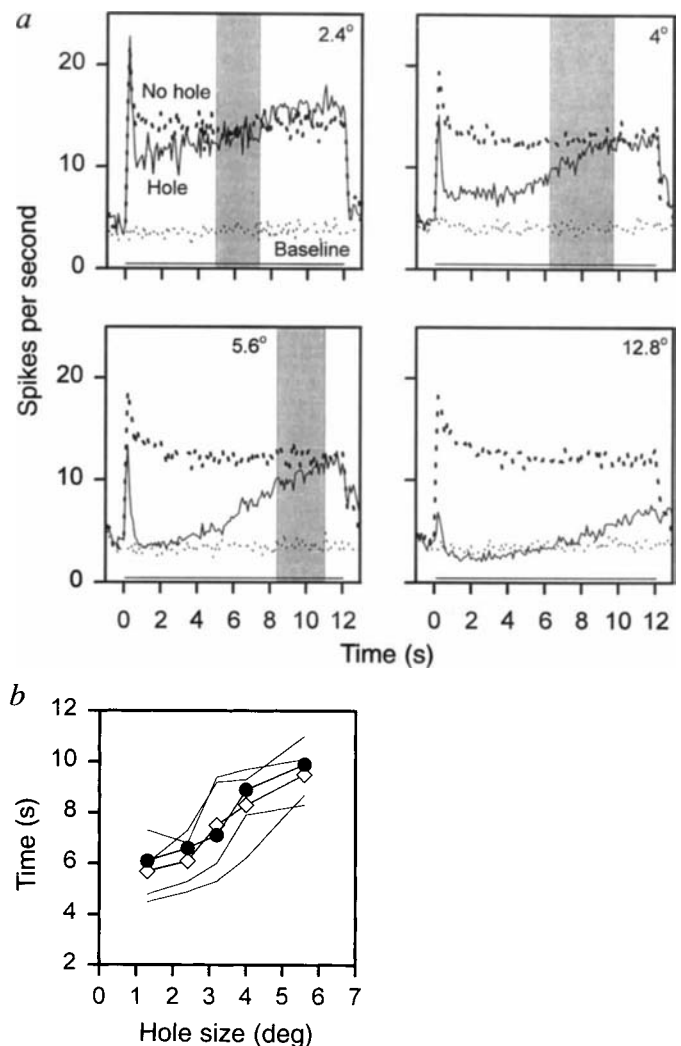


FIG. 4 The role of surround-stimulus characteristics and receptive field expansion in perceptual filling-in and climbing activity. *a*, Single V3 cell showing strong climbing activity when its receptive field was covered by a hole of 4° in a large dynamic texture (thick line), and absence of climbing activity when its receptive field was surrounded by a narrow frame of the same, dynamic texture (thin line). Similar results were obtained in six additional V3 cells with climbing activity; four of the cells tested with the narrow frame showed the same onset transient immediately followed by a low level of activity as they did with the large texture field. *b*, Single V3 cell showing strong climbing activity to a texture with a 4° hole over its receptive field (thick line), and response to the same stimulus flickered at 1.9 Hz (thin line). The flickering strongly suppressed the climbing activity. Instead, the cell gave transient responses time-locked to the flicker. This same flickering stimulus was tested on three other V3 cells that showed climbing activity, and in all cases the flickering suppressed or eliminated the climbing activity. Both the narrow texture frame and the flickering stimulus prevented filling-in in human subjects. *c*, Responses of a single V3 cell to a bar swept just outside the receptive field borders before and after presentation of a texture stimulus (4° hole) that caused climbing activity. One end of the 1.5° -long bar touched the receptive field edge (position 1 in *d*). The two bold line segments on the left of the abscissa indicate the forwards and the backwards motion of the bar before the presentation of the textured stimulus (thin line on abscissa); the two bold line segments on the right indicate the bar's forwards and backwards motion after the presentation of the texture stimulus. *d*, Average receptive field profiles of four cells with climbing activity. Receptive fields were mapped with moving bars just before (open bars) and after (solid bars) presentation of the texture with hole. The bars were swept through areas occupied at a different time by the texture (position 1 and 5) or through the hole and receptive field (positions 2–4). The dotted line indicates average spontaneous activity. Error bars indicate standard errors. There was no evidence for receptive field expansion.

between excitation and inhibition which leads to climbing activity and perceptual filling-in. In permanent scotomas, such as the blind spot, inhibition from the surround may be adapted in a more permanent manner so that filling-in is instantaneous^{7,18,22}. □

Received 8 May; accepted 21 August 1995.

1. Bender, M. B. & Teuber, H. L. *Archs Neurol Psychiat.*, Chicago **55**, 627–658 (1946).
2. Grossberg, S. *Percept. Psychophys.* **55**, 48–120 (1994).
3. Ramachandran, V. S. & Gregory, R. L. *Nature* **350**, 699–702 (1991).
4. Ramachandran, V. S., Gregory, R. L. & Aiken, W. *Vision Res.* **33**, 717–721 (1993).
5. Spillmann, L. & Kurtenbach, A. *Vision Res.* **32**, 1941–1946 (1992).
6. Pettet, M. W. & Gilbert, C. D. *Proc. natn. Acad. Sci. U.S.A.* **89**, 8366–8370 (1992).
7. Komatsu, H. & Murakami, I. *Vis. Neurosci.* **11**, 1103–1113 (1994).
8. Ohzawa, I., Sclar, G. & Freeman, R. D. *J. Neurophysiol.* **42**, 833–849 (1985).
9. Knierim, J. J. & Van Essen, D. C. *J. Neurophysiol.* **67**, 961–980 (1992).
10. Maffei, L. & Fiorentini, A. *Vision Res.* **16**, 1131–1139 (1976).
11. Nelson, J. I. & Frost, B. J. *Brain Res.* **139**, 359–365 (1978).
12. Nelson, J. I. & Frost, B. J. *Expl Brain Res.* **61**, 54–61 (1985).
13. Zipser, K., Lee, T. S., Lamme, V. A. F. & Schiller, P. H. *Soc. Neurosci. Abstr.* **20**, 1477 (1994).
14. Yoshioka, T., Levitt, J. B. & Lund, J. S. *J. Neurosci.* **12**, 2785–2802 (1992).
15. Levitt, J. B., Yoshioka, T. & Lund, J. S. *J. comp. Neurol.* **342**, 551–570 (1994).
16. McGuire, B. A., Gilbert, C. D., Rivilin, P. K. & Wiesel, T. N. *J. comp. Neurol.* **305**, 370–392 (1991).
17. Distler, C., Boussaoud, D., Desimone, R. & Ungerleider, L. G. *J. comp. Neurol.* **334**, 125–150 (1993).
18. Fiorani, M., Rosa, M. G. P., Gattass, R. & Rocha-Miranda, C. E. *Proc. natn. Acad. Sci. U.S.A.* **89**, 8547–8551 (1992).
19. Gattass, R., Gross, C. G. & Sandell, J. H. *J. comp. Neurol.* **201**, 519–539 (1981).
20. Gattass, R., Sousa, A. P. B. & Gross, C. G. *J. Neurosci.* **8**, 1831–1845 (1988).
21. Miller, E. K., Li, L. & Desimone, R. *J. Neurosci.* **13**, 1460–1478 (1993).
22. Das, A. & Gilbert, C. D. *Nature* **375**, 780–784 (1995).

ACKNOWLEDGEMENTS. We thank M. Mishkin for support, and L. Chelazzi, B. Jagadeesh, A. Karni and A. Lueschow for valuable discussions.

Differential properties of two gap junctional pathways made by AII amacrine cells

Stephen L. Mills & Stephen C. Massey

Department of Ophthalmology and Visual Science,
University of Texas at Houston, Health Science Center, Houston,
Texas 77030, USA

THE retina is sensitive to light stimuli varying over more than 12 log units in intensity. It accomplishes this, in part, by switching between rod-dominated circuits designed for maximum utilization of scarce photons and cone circuits designed for greater acuity. Rod signals are integrated into the cone pathways through AII amacrine cells, which are connected by gap junctions both to other AII amacrine cells and to cone bipolar cells. To determine the relative permeabilities of the two junctional pathways, we have measured the distribution of biotinylated tracers across this heterologous cell assembly after injecting a single AII amacrine cell. We found that neurobiotin (relative molecular mass, 286) passed easily through both types of gap junctions, but that biotin-X cadaverine (relative molecular mass, 442) passed through AII/bipolar cell gap junctions poorly compared to AII/AII gap junctions. Thus, the AII/bipolar cell channel has a lower permeability to large molecules than does the AII/AII amacrine cell channel. The two pathways are also regulated differently. Dopamine and cyclic AMP agonists, known to diminish AII–AII coupling¹, did not change the relative labelling intensity of AII to bipolar cells. However, nitric oxide and cGMP agonists selectively reduced labelling in bipolar cells relative to AII amacrine cells, perhaps by acting at the bipolar side of this gap junction. This suggests that increased cGMP controls the network switching between rod and cone pathways associated with light adaptation.

In the mammalian retina, ganglion cells, the final output of the retina, receive input from both rod and cone pathways.

Cones contact ON and OFF cone bipolar cells which synapse directly with ganglion cells. In contrast, rods synapse onto a single type of rod bipolar cell, which does not contact ganglion cells directly, but relays its signal through the AII, or rod amacrine cell^{2–4}. Rod input into the ON pathway occurs through gap junctions between AII amacrine cells and ON cone bipolar cells. The AII also makes gap junctions with other AII amacrine cells (Fig. 1). Tracer coupling across both types of junctions is found with neurobiotin, but not Lucifer yellow¹.

Coupling is most commonly found between similar cells of the same type (homologous). Channels may be formed from pairings of a single type of connexon (homotypic)^{5,6} or two different connexon types (heterotypic)^{7–10}. Total permeability between two cells is the product of the number of open channels times the single-channel permeability. If two gap junctions contain a single connexon type, differences in total permeability to a given tracer can only reflect differences in number of open connexons. Under these conditions, the relative partitioning of dye will be independent of tracer size. Different connexons, on the other hand, can selectively favour passage of one tracer over another. Channels with a small pore diameter will select against larger-diameter tracers^{11–13}. Consequently, if the relative distribution of tracer across two pathways changes with different-sized tracers, then their gap junctions must contain channels with fundamentally different permeabilities. Neurobiotin and biotin-X cadaverine both bear a single positive charge and have similar detection efficiency, as seen by dotblots. It is therefore likely that differences in permeability ratio reflect differences in connexon type, although post-translational modifications of a single type cannot be excluded.

Figure 2 shows examples of AII and bipolar cell staining following injection of biotin-X cadaverine (BXC) or neurobiotin (NBT) into a single AII amacrine cell. Although the number of amacrine cells stained is similar (Fig. 2a,d), far fewer bipolar cells (Fig. 2b,e) are labelled with biotin-X cadaverine (M_r 442) than with neurobiotin (M_r 286) (Fig. 2c,f). Failure of Lucifer yellow (M_r 443) to pass either channel most likely reflects charge selectivity of the channels¹⁴. A quantitative representation of neurobiotin labelling is depicted in Fig. 3a, which shows the relative staining intensity of AII amacrine cells (circles) and cone bipolar

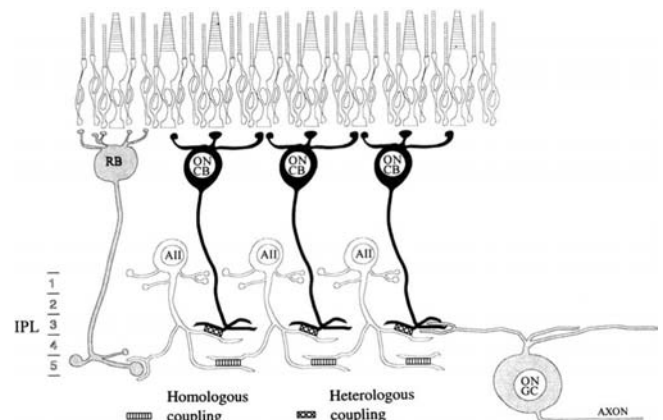


FIG. 1 The ON pathways in mammalian retina. Rods and cones (top) synapse on rod bipolar or cone bipolar cells. ON cone bipolars synapse directly upon ganglion cells (ON GC), rod bipolars (grey, RB) synapse only upon the AII amacrine cell, which transmits rod information to both OFF (not shown) and ON cone bipolar cells (ON CB). AII amacrine cells make gap junctions with ON cone bipolar cells in sublaminae 3 and 4 and also with other AII amacrine cells. The heterologous AII-to-bipolar cell connexons have a lower permeability to large ions than the homologous AII–AII connexons, and are gated by NO and cGMP. The homologous gap junctions are gated by dopamine (DA) and cAMP. IPL, inner plexiform layer.

cells (squares) as a function of distance from the site of injection for a control (no drug) retina. It demonstrates that in control retina, the mean intensity of bipolar cells of diameter $\geq 100 \mu\text{m}$ from the site of injection is $75 \pm 4\%$ (s.e.m.) of that of neighbouring All amacrine cells.

Thus, the ratio of the staining intensity for bipolar cells/amacrine cells (b/a ratio) at comparable distances from the site of injection can be used to compare tracers or pharmacological treatments. This method removes much of the variance attributable to the amount of dye injected or to the diffusion

FIG. 2 Sample injections of tracer into single All amacrine cells. The diagnostic processes of the All amacrine cells are easily seen at focal levels not shown. Some leakage is seen around each injection site. The invariant pattern of cells stained is never found with extracellular ejection and excludes uptake artefacts. Scale bar applies to all panels. **a–c**, Neurobiotin injection: **a**, Focus is on the somas of All amacrine cells. **b**, Focus is on the coupled mosaic of ON cone bipolar cells. **c**, All amacrine cells from **a** and cone bipolar cells from **b** are drawn superimposed. Dashed line shows the radius of neurobiotin-coupled bipolar cells. Neurobiotin injection stained 85 All amacrine cells and 71 cone bipolar cells (ratio, 0.84). **d–f**, Biotin-X cadaverine injection produced intense staining of All somas (**d**), but much less staining of bipolar somas (arrowheads, **e**). **f**, All amacrine cells from **d** and cone bipolar cells from **e** are drawn superimposed. Dashed line shows the radius of biotin-X-cadaverine-coupled bipolar cells. Biotin-X cadaverine injection stained 73 All amacrine cells and 13 cone bipolar cells (ratio, 0.18).

METHODS. All amacrine cells labelled with Nuclear yellow¹ were filled²⁹ (+1 nA for 4 min at 3 Hz) with 1 or 4% neurobiotin or 1% biotin-X cadaverine. Dyes were allowed to diffuse for at least 40 min post-injection. Stained patches increased moderately in size with increased perfusion time. Although the intensities of the coupled cells changed relative to the injected cell with perfusion time, the b/a ratio rapidly declined spatially to an asymptote about $100 \mu\text{m}$ from the injected cell which was stable over intervals even longer than those measured. All drugs were added to the perfusate. Following tissue fixation (in 4% paraformaldehyde for 1 h), cells were visualized with 1:200 streptavidin-CY3. Drugs were purchased from Sigma, except Sp-8-pCPT-cGMPs (BioLog) and SNAP (RBI).

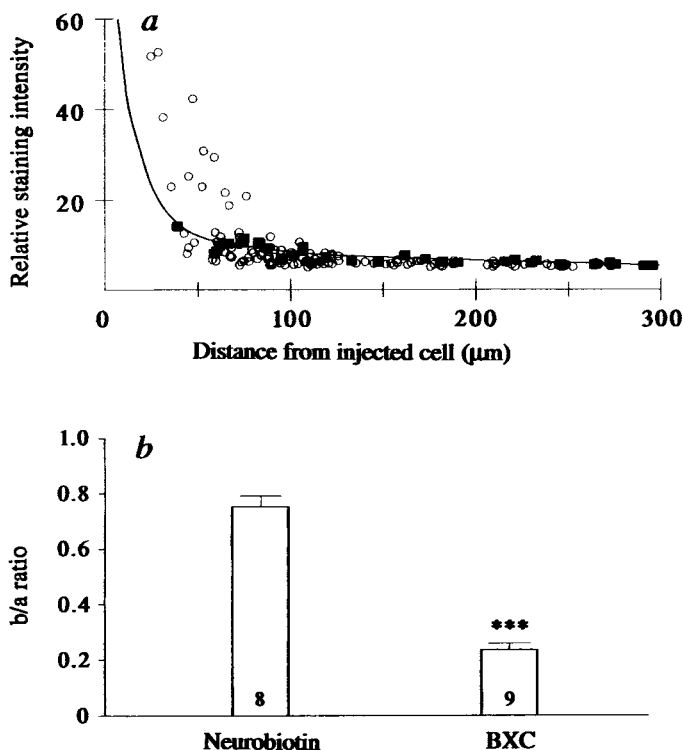
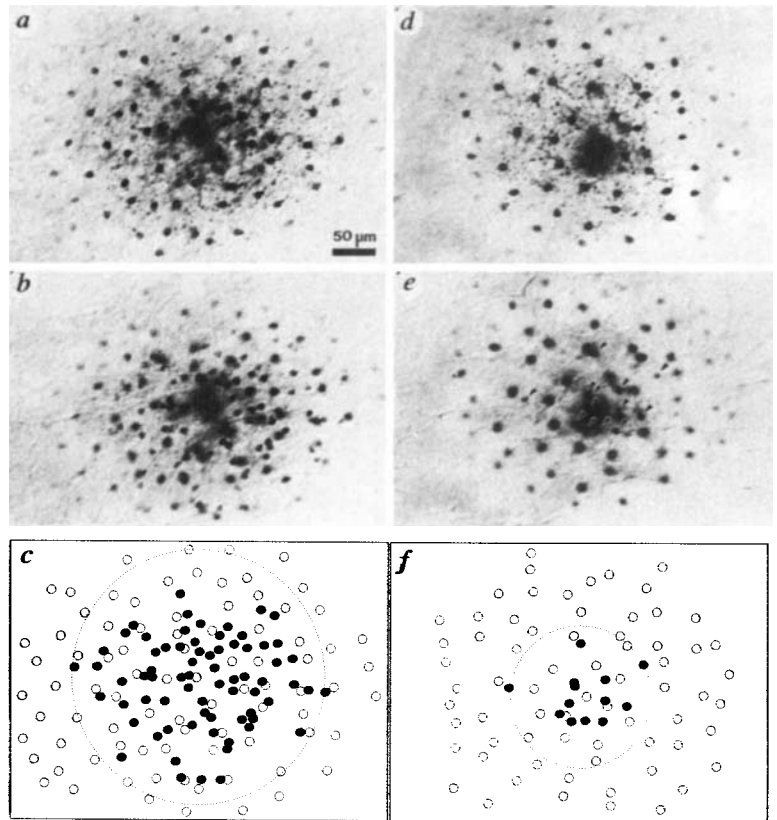


FIG. 3 **a**, Staining intensity of cells as a function of distance from a cell injected with 4% neurobiotin. The intensity relative to background staining is plotted for All amacrine cells (filled squares, connected by smooth line) and ON cone bipolar cells (circles) for a control injection. Cone bipolar cells are initially brighter than equidistant coupled amacrine cells, but fall off more rapidly. The apparent incongruity of a poorly coupled cell being more intensely stained than a neighbouring more well-coupled cell occurs because all cells near the injected site rapidly acquire dye, but only the better-coupled All amacrine cells are able to release it to other cells over the time of the experiment. This behaviour is easily reproduced with model differential equations. The spatial asymptote reflects relative permeabilities of the two pathways. **b**, Relative bipolar/amacrine cell staining (mean $\pm$ 2 s.e.m.) in the region from 100 – $200 \mu\text{m}$ from the injected cell. With neurobiotin as the tracer, bipolar cells average 75% of the intensity of neighbouring All amacrine cells as in **a**. The b/a ratio averages only 23% following biotin-X cadaverine (BXC) injections. The staining intensity of each bipolar cell was divided by the intensity of the nearest All amacrine cell. All such normalized bipolar cells in the range of 100 – $200 \mu\text{m}$ from the cell injected with tracer were averaged. Means and standard error bars in Figs 3 and 4 were calculated from the means of these individual injections, and significance levels represent results of *t*-tests between control and experimental treatments. ***, $P < 0.001$. Number of injections appears inside bars. Images were acquired using ImagePro. Relative staining intensities of cells were measured in Adobe Photoshop by centring a 5×5 pixel grid on each cell. This digitized intensity was converted to a calibrated brightness value relative to the background, set to 1.0.

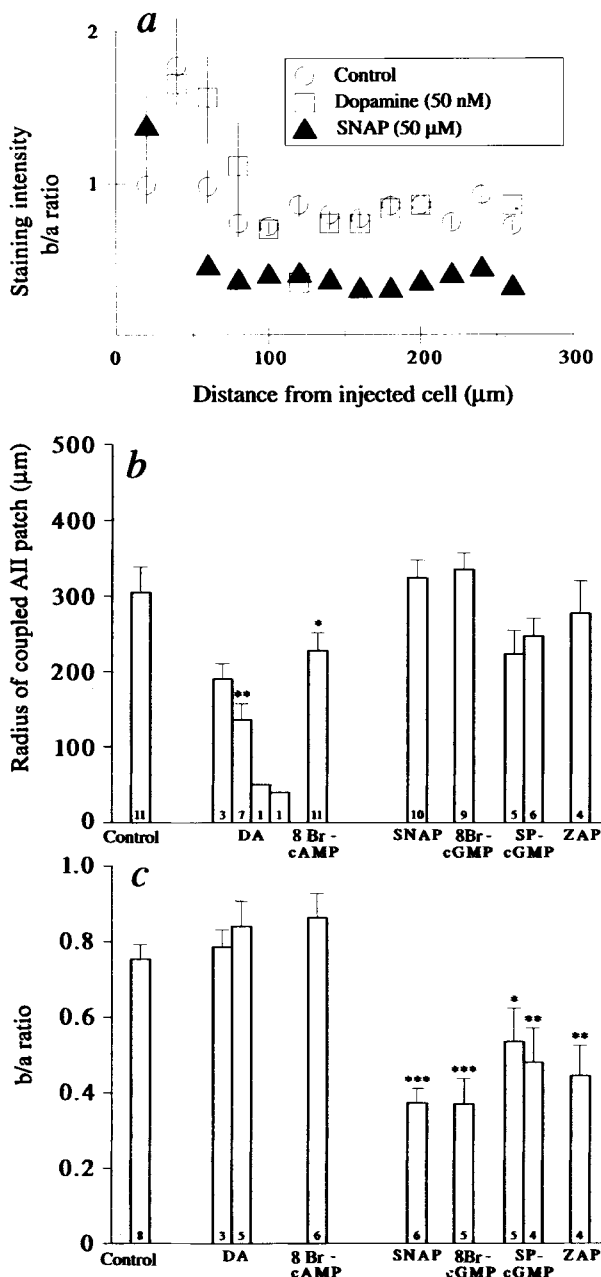


FIG. 4 *a*, The average staining intensity of cone bipolars relative to their nearest AII amacrine cell. Relative *b/a* intensities exceeding 2.0 are not shown and occur only at distances of less than 50 μm. Error bars (± 1 s.e.m.) represent *b/a* ratios across all bipolar cells, injections and retinas. The cGMP-stimulator SNAP reduced staining in bipolar cells; dopamine had no effect at concentrations that significantly reduce AII–AII coupling. Concentrations of dopamine or 8-bromo-cAMP that strongly reduced coupling precluded comparisons at more than 50 μm from the injected cell and produced ratios of *b/a* staining over the peri-injection region that were similar to controls. *b*, *c*, Comparison of neurobiotin coupling between AII amacrine cells (*b*) and from AII amacrine cells to bipolar cells (*c*). All conditions ($n > 1$) represent multiple retinas. *b*, The radius of the coupled mosaic of AII amacrine cells. Agents that increase cAMP reduce coupling; agents that increase cGMP do not. Drug concentrations (number of injections in parentheses) were, from left to right, control (11); dopamine: 10 nM (3), 50 nM (7), 100 nM (1), 1 μM (1); 8-bromo-cAMP: 125 μM (11); SNAP: 50 μM (10); 8-bromo-cGMP: 400 μM (9); Sp-8-pCPT-cGMPs: 100 nM (5), 1 μM (6); and zaprinast (ZAP): 25 μM (4). High concentrations of dopamine were not investigated further, because patches were smaller than the 100–200 μm test range. *c*, The average relative bipolar/AII amacrine cell staining intensity. Drug concentrations (number of injections) were control (8); dopamine: 10 nM (3), 50 nM (5); 8-bromo-cAMP: 125 μM (6); SNAP: 50 μM (6); 8-bromo-cGMP: 400 μM (5); Sp-8-pCPT-cGMPs: 100 nM (5), 1 μM (5), 1 μM (4); and zaprinast: 25 μM (4). Dopamine concentrations greater than 50 nM produced stained patches too small to analyse. Agents that produce or mimic rises in intracellular cGMP reduced staining in bipolar cells, including SNAP, zaprinast, and two membrane-permeant cGMP analogues, Sp-8-pCPT-cGMPs and 8-bromo-cGMP. Dopamine and 8-bromo-cAMP did not alter the ratio of staining intensities from those of controls. Significance levels: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

time. Although this ratio will approach unity with a long incubation time, it is stable over the intervals measured in these experiments. As the initial partitioning of dye will match the relative permeabilities of the two pathways, this stability of the *b/a* ratio suggests that it is a reliable estimate of the relative permeabilities under the present set of conditions. Figure 3*b* shows the *b/a* ratio for neurobiotin and biotin-X cadaverine, averaged over the region from 100–200 μm away from the injected cell. Although biotin-X cadaverine, the large dye molecule, moved less readily through AII/AII gap junctions, as seen by decrease in average size of the stained patch (not shown), the dramatic decrease in *b/a* ratio (Fig. 3*b*) shows that permeability of the AII/bipolar cell junctions decreased far more. This suggests that the channels between AII amacrine and bipolar cells have a smaller pore size, and hence a different connexon type in at least one hemichannel. It is unclear whether the AII amacrine cell expresses two types of connexin, with both pathways homotypic but different, or

expresses only a single type, with the AII/bipolar cell channel being heterotypic.

The AII amacrine cell is an obligatory part of the high-gain rod pathway⁴; its output into the ON pathway is through gap junctions with ON cone bipolar cells. Dopamine, which is involved in the transition from rod to cone vision by many mechanisms^{15–17}, dramatically decreases the permeability of gap junctions connecting AII amacrine cells as a result of increasing the cAMP concentration. This is seen as reductions in patch size, taken as either the number¹ or radius (Fig. 4*b*) of somas whose staining is visually detectable. As the cone bipolars may only receive tracer from the AII amacrine cells with whom they make gap junctions, any reduction in size of the group of coupled amacrine cells concomitantly reduces the number of stained bipolar cells. Thus, changes in coupling between the AII amacrine cell and ON bipolar cells are not easily detected by measuring the size of the stained patch. A more sensitive measure is quantification of relative staining intensity of AII amacrine and

cone bipolar cells at comparable distances from the site of injection.

After neurobiotin injection in dopamine-treated retina, even the smallest coupled groups retain bipolar cell labelling, and staining intensity of cone bipolars relative to amacrine cells (Fig. 4a, squares) is similar to untreated controls (Fig. 4a, circles). This holds true even for concentrations that significantly reduce the size of the stained patch (Fig. 4b, c). Similar results were obtained with 8-Br-cAMP (Fig. 4b, c). However, in tissue incubated in the nitric oxide (NO) donor *S*-nitroso-*N*-acetylpenicillamine (SNAP) (Fig. 4a, triangles), which elevates cGMP in ON cone bipolar cells^{18,19}, coupled bipolars averaged only about 40% of the staining intensity of AII amacrine cells. The effect of SNAP was similar to that of the membrane-permeant cGMP analogues 8-Br-cGMP and Sp-8-pCPT-cGMPs (Fig. 4c). The cGMP-phosphodiesterase inhibitor zaprinast (25 μ M) also selectively reduced bipolar staining. None of the cGMP-elevating compounds significantly changed the radius of the labelled patch of AII amacrine cells (Fig. 4b). Our results confirm that coupling between members of the AII amacrine cell class is modulated by dopamine through cAMP, and show by the use of four different substances that the junctional resistance between AII amacrine cells and ON bipolar cells is modulated through increases in cGMP concentration.

The channels between AII amacrine cells therefore differ in two fashions from those between AII amacrine cells and cone bipolar cells. The AII/bipolar cell channel is gated not by cAMP, but by cGMP, and differential permeability to large molecules shows that it contains a different connexin type in at least one hemichannel. The AII/AII channels presumably subserve changes in receptive field size with ambient light, which would improve signal-to-noise ratio by pooling information from many cells^{1-4,20}. The AII/bipolar cell channels, on the other hand, may uncouple dark and light pathways altogether during light adaptation. In dim light, the AII-to-bipolar cell gap junctions must be open to relay the rod bipolar signal to ON cone bipolar cells. In daylight, this pathway would only decrease acuity and dissipate signals passed directly from cones to cone bipolar cells. It would therefore be useful to close these junctions with increasing ambient light.

The mechanism by which this occurs appears to be by increasing the level of cGMP. ON cone bipolar cells show a dramatic and selective increase in levels of cGMP following exposure to NO donors^{18,19}. As NO may be released in the inner plexiform layer²¹ and has been shown to increase the rate of activation of cGMP-gated channels in retinal bipolar cells²², it may also regulate cGMP-gated gap junctions. If NO is released by light stimulation¹⁸, it could amplify the bipolar cell light signal by means of soluble guanylate cyclase²² and simultaneously reduce a potential shunt to the AII amacrine cell. Alternatively, it may be that an increase in cGMP in ON cone bipolar pathways, which follows reduced stimulation of the metabotropic mGluR6 receptor with increasing levels of light²³⁻²⁸, can by itself close a pathway that would dissipate daylight signals to a rod pathway of declining utility. In either case, cGMP modulation of AII/bipolar cell gap junctions plays a role in the pathway switching that underlies light adaptation. □

14. Veenstra, R. D., Wang, H.-Z., Beyer, E. C. & Brink, P. R. *Circ. Res.* **75**, 483-490 (1994).
15. Törk, I. & Stone, J. *Brain Res.* **189**, 261-273 (1979).
16. Voigt, T. & Wässle, H. *J. Neurosci.* **7**, 4115-4128 (1987).
17. Witkovsky, P. & Dearry, A. *Progr. Ret. Res.* **11**, 247-292 (1991).
18. Koistinaho, J., Swanson, R. A., De Vente, J. & Sagar, S. M. *Neuroscience* **57**, 587-597 (1993).
19. Massey, S. C., Mills, S. L. & De Vente, J. *Invest. Ophthalmol. Vis. Sci.* (suppl.) **34**, 1382 (1993).
20. Vaney, D. I. *Progr. Ret. Res.* **13**, 301-355 (1994).
21. Yamamoto, R., Bredt, D. S., Snyder, S. H. & Stone, R. A. *Neuroscience* **54**, 189-200 (1993).
22. Shiels, R. & Falk, G. *Neuroreport* **3**, 845-848 (1992).
23. Slaughter, M. M. & Miller, R. F. *Science* **211**, 182-185 (1981).
24. Shiels, R. & Falk, G. *Proc. R. Soc. Lond. B* **242**, 91-94 (1990).
25. Nawy, S. & Jahr, C. E. *Nature* **346**, 269-271 (1990).
26. Yamashita, M. & Wässle, H. *J. Neurosci.* **11**, 2372-2382 (1991).
27. Nomura, A. et al. *Cell* **77**, 361-369 (1994).
28. Tian, N. & Slaughter, M. M. *J. Neurophysiol.* **71**, 2258-2268 (1994).
29. Mills, S. L. & Massey, S. C. *J. comp. Neurol.* **304**, 491-501 (1991).

ACKNOWLEDGEMENTS. This research was supported by NIH grants to S.L.M. and S.C.M., an NIH core grant to University of Texas at Houston, and RPB funds to the Department of Ophthalmology and Visual Science. We thank W. S. Sunny Liu for her contribution to this project and D. Marshak for many useful comments on this manuscript.

Unitary anion currents through phospholemman channel molecules

J. Randall Moorman^{*†}, Stephen J. Ackerman^{*},
Gopal C. Kowdley^{*}, M. Pamela Griffin[‡],
J. Paul Mounsey^{*¶}, Zhenhui Chen[§],
Steven E. Cala[§], Jeffrey J. O'Brian^{||}, Gabor Szabo[†]
& Larry R. Jones[§]

Departments of ^{*} Internal Medicine (Cardiovascular Division),
[†] Molecular Physiology and Biological Physics, and [‡] Pediatrics
(Neonatology Division), University of Virginia Health Sciences Center,
Charlottesville, Virginia 22908, USA
[§] Department of Internal Medicine, Krannert Institute of Cardiology,
Indiana University School of Medicine, Indianapolis,
Indiana 46202, USA
^{||} DuPont Merck Pharmaceutical Company, Wilmington,
Delaware 19898, USA

PHOSPHOLEMMAN (PLM) is a 72-amino-acid peptide with a single transmembrane domain¹⁻³, the expression of which induces chloride currents in *Xenopus* oocytes⁴⁻⁶. It has remained unknown whether PLM is an ion channel or acts as a channel regulator. Here we show, by measuring unitary anion currents across planar phospholipid bilayers to which immunoaffinity-purified recombinant PLM was added, that it does indeed form ion channels. Excised patches of oocytes expressing PLM had similar currents. Of the ions tested, the sulphonic amino acid taurine was the most permeant, and expression of PLM increased fluxes of radiolabelled taurine in oocytes. Phospholemman is the smallest protein in cell membranes known to form an ion channel and the taurine selectivity suggests that it is involved in cell volume regulation.

The effect of adding purified phospholemman (PLM) to lipid bilayers is shown in Fig. 1. The appearance of multiple ion conductance levels from 250 to 3,200 pS at 50 mV in symmetric 150 mM NaCl at room temperature is shown in Fig. 1a. We suggest that the increasing conductance resulted from incorporation of PLM channel molecules in the lipid bilayer. Once open, the channels rarely closed over potentials between -75 and +75 mV. Current-voltage relations measured during 10-s voltage ramps over this range are shown in Fig. 1b-e, with 200 mM KCl on the side of the reference electrode and 50 mM KCl on the other. Uncorrected data are shown in Fig. 1b, c; the results in b were obtained before addition of protein, and indicate a capacitive transient and a leak conductance of 10 pS. The cur-

Received 11 July; accepted 30 August 1995.

1. Hampson, E. C. G. M., Vaney, D. I. & Weiler, R. J. *Neurosci.* **12**, 4911-4922 (1992).
2. Kolb, H. & Famiglietti, E. V. *Science* **186**, 47-49 (1974).
3. Strettoi, E., Raviola, E. & Dacheux, R. F. *J. comp. Neurol.* **325**, 152-168 (1992).
4. Sterling, P. A. *Rev. Neurosci.* **6**, 149-185 (1983).
5. Bennett, M. V. L. et al. *Neuron* **6**, 305-320 (1991).
6. Beyer, E. C., Paul, D. L. & Goodenough, D. A. *J. Membr. Biol.* **116**, 187-194 (1990).
7. Rae, J. L. *Curr. Top. Eye Res.* **1**, 37-90 (1979).
8. Swenson, K. L., Jordan, J. R., Beyer, E. C. & Paul, D. L. *Cell* **57**, 145-155 (1989).
9. Werner, R., Levine, E., Rabadan-Diehl, C. & Dahl, G. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5380-5384 (1989).
10. Bruzzone, R., White, T. W. & Paul, D. L. *J. Cell Sci.* **107**, 955-967 (1994).
11. Flagg-Newton, J., Simpson, I. & Loewenstein, W. R. *Science* **205**, 406-407 (1979).
12. Brink, P. & Dewey, M. M. *Nature* **285**, 101-102 (1980).
13. Zimmerman, A. L. & Rose, B. J. *Membr. Biol.* **84**, 269-283 (1985).

^{*} Present address: Department of Cardiovascular Medicine, University of Birmingham, Birmingham B15 2TT, UK.

rent shown in Fig. 1c was obtained a few minutes after adding purified recombinant PLM. There is an abrupt initiation of a conductance at about -50 mV. The results in Fig. 1d, e were obtained after subtracting the mean of several traces obtained before the addition of protein. The corrected trace with channel activity (Fig. 1e) allows calculation of conductance (700 pS) and E_{rev} (17 mV). The positive value of E_{rev} establishes anion-selectivity; at 17 mV, $P_K/P_{Cl} \approx 0.3$. In separate experiments, we also found $P_{Na}/P_{Cl} \approx 0.3$ ($n = 10$). A frequency histogram of the value of the initial conductance for 62 experiments in asymmetric KCl conditions was described by a single gaussian function with mean 645 pS and s.d. 168 pS. This distribution supports the idea that each of these well-defined initial events consist of the insertion of a single entity. A frequency histogram of the voltage at which the initial conductance appeared showed a uniform distribution, suggesting that channel incorporation was not voltage dependent.

To demonstrate further that PLM was the molecule responsible for these currents, we tested the effect of affinity-purified antibodies to the 15 amino-terminal amino acids of PLM as well as a monoclonal antibody to an epitope localized to the N-terminal half of the molecule. The interaction of antibodies with the channel-forming structure (Fig. 2), causing complete block or flickering block, suggests that the antibody interacts with the channel pore. We have observed similar results in 55 bilayer experiments, and we interpret the results to mean that PLM molecules form the channel.

We measured the relative permeability of the PLM channel for several different anions (Fig. 2e and Table 1). For halides,

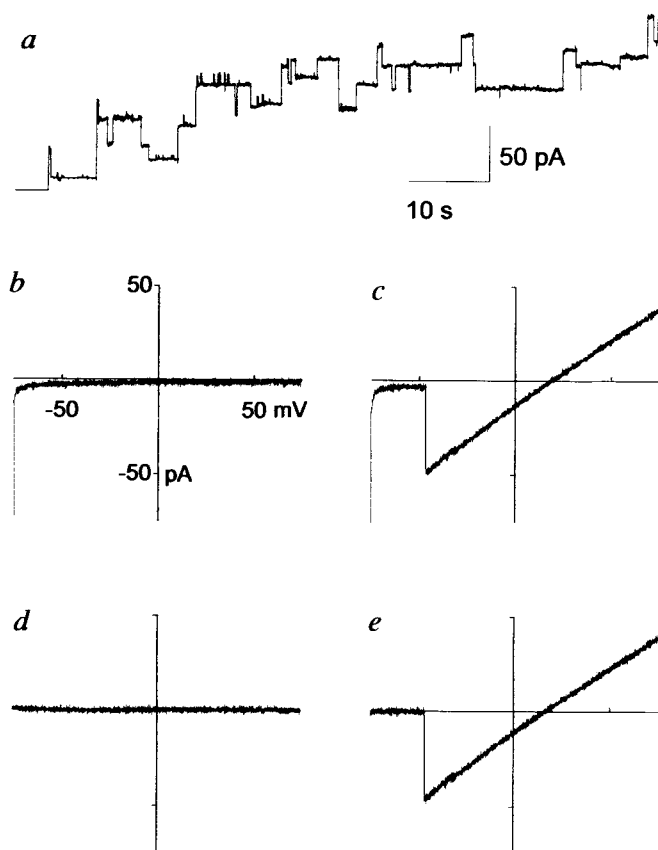
the selectivity sequence $Cl^- > Br^- > F^-$ was found. This is a type IV or V sequence, and indicates an intermediate-strength anionic site⁷. Given the dimensions of benzenesulphonate, the largest permeant anion tested, the lumen of the channel must be larger than 8×9 Å. Although the pore is relatively large, there must be features of the pore that allow for selectivity, as the order of permeabilities of the anions does not match the order of their free solution mobilities as estimated by their limiting equivalent conductivities (Table 1). For instance, the permeability of NO_3^- is greater than that of Cl^- even though its equivalent conductivity is not. Thus the PLM channel does not select solely on the basis of hydrated anion size: other properties of the anion also dictate permeation.

Taurine was by far the most permeant ion tested, and substitution of the amino group by hydroxyl (isethionate) or methyl (propanesulphonic acid) groups, or by Br (2-bromoethanesulphonic acid) drastically reduced permeation (Table 1). Taurine forms zwitterions, and only 4% of taurine molecules are anionic at pH 7.4. The currents we record are thus carried by a minority of the taurine molecules present. This suggests that the taurine flux may be larger than the ion-current measurement predicts, and that pH may modulate the channel's conductance.

To test the idea that PLM molecules make similar channels in *Xenopus* oocytes, we measured currents across membrane patches from oocytes expressing PLM. Currents across an excised inside-out patch during a 10-s voltage ramp in a fourfold concentration gradient of KCl are shown in Fig. 3a. Just after patch excision, conductance was very low (trace 1, 60 pS, E_{rev} 0 mV). A large, anion-selective conductance (trace 2, 6.6 nS, E_{rev}

FIG. 1 Ionic currents through PLM molecules reconstituted in planar lipid bilayers. a, Incorporation of PLM channels in symmetric 150 mM NaCl at 50 mV. b–e, Current–voltage relations measured during a voltage ramp before (b, d) and after (c, e) addition of purified PLM; b, c, uncorrected; d, e, obtained after subtracting the mean of several traces obtained before protein was added. There is an abrupt initiation of a conductance at about -50 mV, and the value of the reversal potential identifies the main charge carrier as Cl^- .

METHODS. Recombinant PLM was immunoaffinity-purified from *Baculovirus*-infected Sf21 insect cells. The coding sequence for PLM (including the region containing the signal peptide sequence) was generated by PCR of base pairs 1–279. The template was the full-length canine cardiac PLM clone in pBluescript³. Primers were constructed with *Nco*I and *Pst*I and sites at the 5' and 3' flanking ends, respectively. This PCR product was then purified and inserted into the *Nco*I and *Pst*I sites of a modified pVL1392 transfer plasmid. After infecting Sf21 cells with the wild-type virus and transfer plasmid, recombinant viral clones are released into the supernatant. Clones containing PLM inserts were isolated by limiting dilution using dot-blot radioimmunoassay to detect expressed PLM. After several rounds of serial dilution and screening pure recombinant viral clones expressing high levels in Sf21 cells were identified. PLM was purified from a crude preparation of insect cell membranes by antibody affinity chromatography using a PLM monoclonal antibody²². The purity of the preparation was verified by SDS-PAGE and protein sequencing, where a single product corresponding to amino acids 1–72 of canine PLM³ was detected. Mass spectroscopy verified the purity of the preparation and showed no evidence of protein phosphorylation (kindly provided by D. Cafiso). Bilayers were formed by painting 1,2-diphytanoylphosphatidylcholine at a concentration of 20 mg ml^{-1} in decane across a 0.45-mm aperture forming the only connection between two Teflon chambers. The front chamber held the reference electrode and (in mM) 50 KCl, 10 HEPES (pH 7.4, KOH). The back chamber, where voltage was applied and current was measured, held a solution of (in mM) 200 KCl, 10 HEPES (pH 7.4, KOH). Salt solutions (1.1 ml) were added to the chambers, membranes were formed, then 1–4 µg purified PLM in 20 mM glycine and 0.1% TritonX-100 (pH 7.2, MOPS) was added to the front chamber. Voltage ramps (10 – 20 mV s^{-1}) were run across the formed bilayer and currents measured using bilayer clamp amplifiers. Currents were amplified and filtered (10 mV pA^{-1} and 100 Hz ; Warner Corporation and Axon Instruments) and digitized (200 Hz , Axon Instruments) and analysed (Origin, Microcal) using conventional means. Formation of a bilayer was monitored by



visual inspection and by measurement of capacitance. The average of traces without channel activity was subtracted from traces with active current. Lines were fit through conductances longer than 200 ms, and reversal potentials were determined from the x axis intercept of these fits.

TABLE 1 Anion permeability of phospholemman channels

(a) Anion: Cl	Limiting G	E_{rev} (mV)	G(nS)	P_{Cl}/P_{Cl}
Taurine (4:127)		-26.37 ± 6.92	0.23 ± 2.53	69.9
NO ₃ (3:87)	71.4	-5.27 ± 5.93	1.73 ± 0.57	1.23
SCN (3:32)	66	-0.34 ± 1.11	1.58 ± 2.1	1.01
Cl (5:40)	76.4	-0.50 ± 0.88	0.71 ± 0.23	1
Benzenesulphonate (3:36)		2.67 ± 0.44	0.96 ± 0.42	0.9
Bicarbonate (4:55)	44.5	3.6 ± 0.84	3.70 ± 2.3	0.87
Benzoate (5:139)	32	5.87 ± 4.23	1.16 ± 0.88	0.8
Lactate (3:37)	38.8	6.13 ± 4.23	1.44 ± 0.70	0.79
Br (2:76)*	78.1	7.12	0.96	0.76
Glutamate (6:119)		7.41 ± 2.3	0.88 ± 1.1	0.75
F (5:106)*	55.4	7.67 ± 3.0	0.86 ± 0.81	0.74
Methylsulphonic acid (9:148)	~39.6	8.42 ± 3.74	1.95 ± 2.29	0.72
H ₂ PO ₄ (7:114)	57	19.74 ± 4.76	0.90 ± 0.66	0.69
Glucuronate (7:174)		10.54 ± 5.04	1.57 ± 2.7	0.66
Isethionate (8:115)		11.2 ± 4.36	1.30 ± 0.97	0.65
Methylsuccinate (3:57)	~58.8	27 ± 2.46	1.62 ± 2.07	0.36

(b) Anion: taurine	E_{rev} (mV)	G(nS)	$P_{taurine}/P_X$
Isethionate (4:47)	-28.5 ± 8.68	0.36 ± 0.17	185
2-bromoethanesulphonate (5:47)	-51.4 ± 9.49	0.30 ± 0.18	96.2
Propanesulphonic acid (4:104)	-19 ± 8.48	0.68 ± 0.70	53.4

Experiments were performed under bi-ionic conditions. In (a), the solutions were 150 mM NaCl in one chamber and 150 mM NaX in the other chamber, where X is the test anion. In (b), the solutions were 150 mM Na taurine and 150 mM NaX. The numbers in parentheses are the number of bilayers tested and the number of current-voltage relations analysed, respectively. For instance, 4 bilayer experiments and a total of 127 current-voltage relations were analysed in Na taurine:NaCl conditions. For each bilayer, the mean E_{rev} and G were calculated; the means from individual bilayers were used to calculate the mean ($\pm$ s.d.) values. 'Limiting G' is the limiting equivalent conductivity in (S per cm)/(equivalent per cm³)²¹. Permeability ratios were determined using the Goldman-Hodgkin-Katz equation from the measured reversal potentials during voltage ramps after subtracting the mean of currents without active channels²¹. All the anions are monovalent except methylsuccinate, which is divalent. Although the concentration of taurine was 150 mM, only 6 mM is anionic at pH 7.4, and this value was used in the calculation of permeability ratios.

* In 150 mM NaBr:NaF, P_{Br}/P_F was 1.1.

24 mV) appeared within 2 min. Trace 3 is current through the broken electrode, and approximates 0 mV. We found similar anion-selective conductances in 22 of 64 patches from oocytes expressing PLM. In uninjected oocytes, where hyperpolarization-activated, non-inactivating chloride currents are sometimes observed⁵, similar patch currents were observed significantly less frequently (5 of 51 patches, $P < 0.05$, chi-squared test).

To show whether PLM can mediate taurine flux across cell membranes, we measured influx and efflux of radiolabelled taurine in oocytes expressing PLM. These oocytes had significantly higher fluxes of taurine than did control oocytes (Fig. 3b), supporting a possible role for PLM in mediating taurine transport.

The high selectivity PLM channels have for taurine suggests a physiological role for the peptide. Heart cells accumulate high concentrations of taurine (30–60 mM), as do neurons and skeletal myocytes⁸, and taurine levels are increased by hyperosmolarity and β -adrenergic activation^{8–10}. It has been shown that cell volume-activated taurine efflux occurs through anion channels in neuronal, kidney and blood cells^{11–14}. We suggest that PLM channels may link signal transduction cascades to taurine flux, allowing precisely adaptive control of cell volume.

We propose that PLM is a member of a new superfamily of membrane proteins, characterized by single transmembrane domains, which function in transmembrane ion flux. Other members of this superfamily are I_{K} (minK) (refs 15–17), Mat-8 (refs 18, 19) and CHIF (ref. 20). All induce ion currents when expressed in oocytes, but it is not known whether the oocyte currents induced by expression of these molecules arise from new channels or from regulated endogenous oocyte channels. Bilayer reconstitution of purified PLM molecules avoids the problems of interpretation inherent in oocyte-expression studies, and allows a direct demonstration of channel formation. The

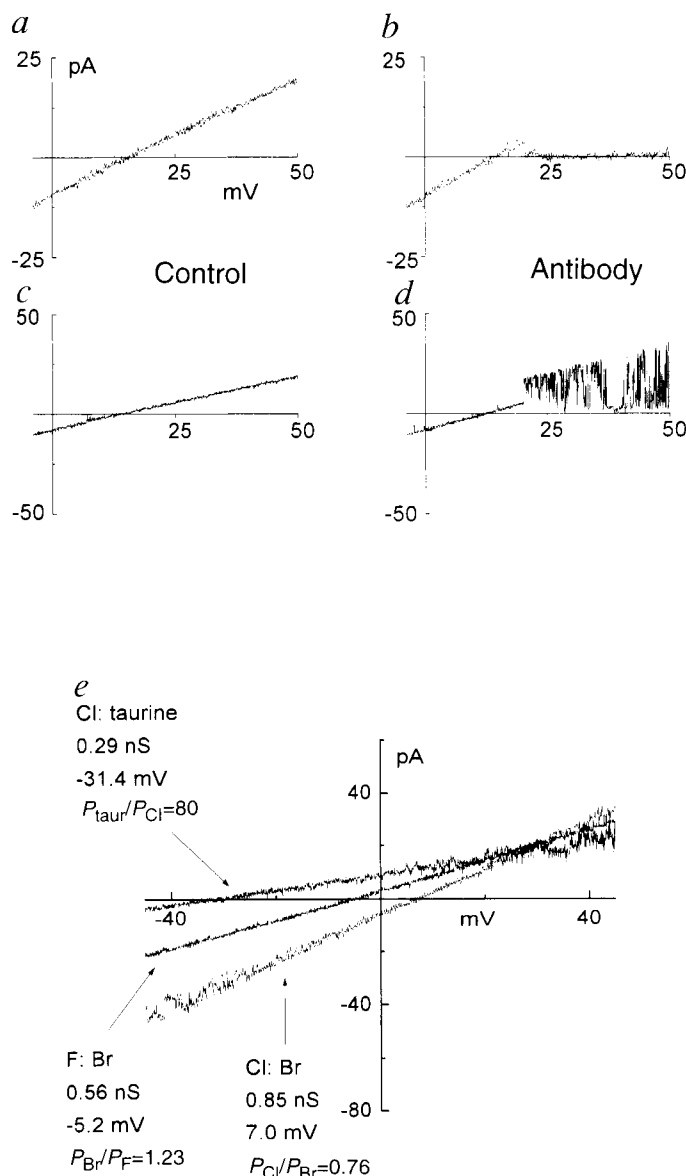
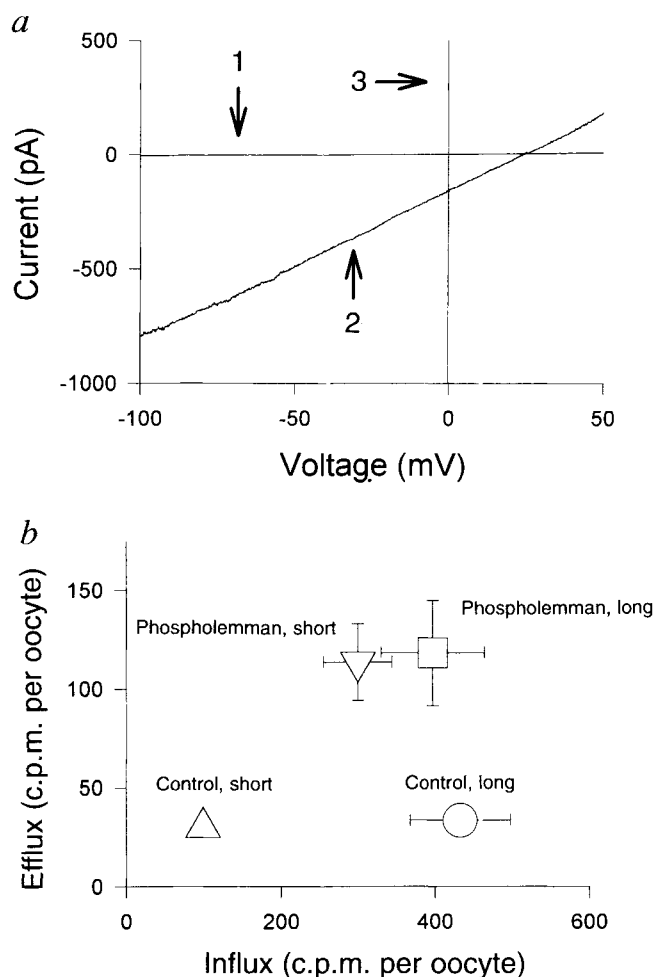


FIG. 2 PLM-specific antibody effects on currents through PLM molecules. a, A bilayer sweep before the addition of polyclonal antibodies raised against the 15 N-terminal residues of PLM. After antibodies were added to the side of the reference electrode (b), the channel closes. In a similar experiment (c, d), monoclonal antibody B8 led to the appearance of 'flickering block'. e, Selectivity of PLM channels. Three current recordings from separate experiments are shown. All were made under 150 mM bi-ionic conditions with Na⁺ as the cation, and conductance, reversal potentials, and permeability ratios are given. Permeability ratios were calculated from the Goldman-Hodgkin-Katz equation. The concentration of anionic taurine in a 150-mM solution at pH 7.4 is 6 mM. The bottom two traces yield the halide selectivity sequence $Cl^- > Br^- > F^-$; the top trace shows the high selectivity for taurine. METHODS. The N-terminal specific antibody was made to a synthetic peptide containing residues 1–15 of PLM with addition of a carboxy-terminal cysteine. Peptides were coupled to thyroglobulin by their cysteine residues, mixed with adjuvant, and injected into rabbits²³. Antibodies were affinity purified from antisera using the peptides coupled to amino-butyl agarose by their cysteine residues²⁴. Monoclonal antibody B8 to PLM was obtained from the spleen cells of an immunized BALB/c mouse²³. The antigen was a partly purified preparation of PLM expressed in Sf21 insect cells. The monoclonal antibody was purified by protein A chromatography of ascites fluid²³. All antibodies used bound to a single band of apparent relative molecular mass (M_r) of 15K in cardiac sarcolemmal vesicles.



pronounced selectivity of reconstituted PLM channels for taurine and the increased taurine fluxes in oocytes expressing PLM demonstrate that PLM is involved in membrane ion transport. □

FIG. 3 *a*, PLM molecules induce similar currents in *Xenopus* oocytes. Currents across an excised patch of an oocyte expressing PLM immediately after excision (trace 1) and 2 min later (trace 2). Trace 3 is current through the broken pipette, near 0 mV. *b*, PLM expression increases taurine fluxes in *Xenopus* oocytes. After incubation in [3 H]taurine for 4 h (short) or 15–18 h (long incubation), oocytes previously injected with PLM RNA or KCl were incubated in a taurine-free hypotonic solution. Radioactivity in the solution and the dried oocytes was then determined. After short incubation, there were significant increases in taurine influx and efflux in oocytes expressing PLM ($P < 0.001$, Mann–Whitney rank sum test). After long incubation, oocytes expressing PLM had significantly higher taurine efflux ($P < 0.001$) but about the same taurine influx (not significant). The simplest interpretation is that PLM expression increased the rates of taurine influx and efflux but did not increase the maximum taurine capacity of the oocytes.

METHODS. Oocytes from *Xenopus laevis* were collected, injected through the follicular layer with 25–70 nl 150 mM KCl, or 150 mM KCl and 2–40 ng of PLM RNA using methods described previously, and incubated for 24–72 h (refs 4, 25). Patch-clamp experiments were performed by conventional techniques. The bath contained (in mM) KCl 200, EGTA 10, HEPES 10; the electrode contained KCl 50, HEPES 10, CaCl_2 3, MgCl_2 2.0 (pH 7.4, KOH). For taurine flux measurements, groups of 5–17 defolliculated oocytes were placed into a physiological saline solution containing 150 mM NaCl, osmolality 280, and 20 $\mu\text{Ci ml}^{-1}$ [3 H]taurine (New England Nuclear). After 4 h (short) or 15–18 h (long incubation), the solution was aspirated, and the oocytes were washed 6 times with 5-ml portions of saline solution without taurine. The oocytes were then placed in modified Barth's solution (MBS), osmolality 220, for 1 h. The MBS was then collected, and the oocytes washed in 3.5 ml fresh MBS, which was also collected. The MBS and the dried oocytes were placed in separate scintillation vials with 5 ml scintillation cocktail, and counts were determined. The counts in the MBS were taken as a measure of taurine efflux, and the sum of the counts in the MBS and in the oocytes was taken as a measure of the taurine influx. Long-incubation experiments were performed with 57 control and 62 PLM-expressing oocytes from 6 frogs; short-incubation experiments were performed with 144 and 99 oocytes from 8 frogs. Data are displayed as the mean of the normalized c.p.m. per oocyte; error bars are 95% confidence intervals.

Follicular dendritic cells and human immunodeficiency virus infectivity

Sonya L. Heath*, J. Grant Tew*, John G. Tew*, Andras K. Szakal† & Gregory F. Burton*†

Department of * Microbiology & Immunology and of † Anatomy, Division of Immunobiology, Box 980678, Virginia Commonwealth University, Richmond, Virginia 23298-0678, USA

LARGE amounts of human immunodeficiency virus (HIV) localize on follicular dendritic cells (FDC) in the follicles of secondary lymphoid tissues following viral infection^{1,2}. During clinical latency, active viral infection occurs primarily at these sites^{3,4}. As HIV on FDC is in the form of immune complexes⁵, some of which may be formed with neutralizing antibody, we investigated whether HIV on FDC is infectious. We report here that HIV on FDC is highly infectious. Furthermore, FDC can convert neutralized HIV into an infectious form even in the presence of a vast excess of neutralizing antibody. Thus FDC may provide a mechanism whereby HIV infection can continue in the presence of neutralizing antibody.

We used an *in vitro* model of the germinal centre, designed to mimic events during clinical latency, to determine whether or not FDC-retained HIV could cause infection. We obtained FDC from the tonsils of non-HIV-infected individuals and incubated

Received 22 May; accepted 24 August 1995.

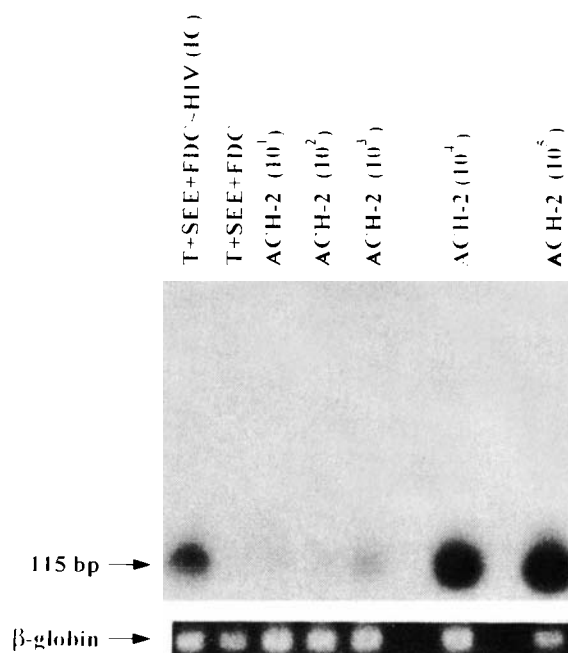
- Presti, C. F., Jones, L. R. & Lindemann, J. P. *J. biol. Chem.* **260**, 3860–3867 (1985).
- Lindemann, J. P. *J. biol. Chem.* **261**, 4860–4867 (1985).
- Palmer, C. J., Scott, B. T. & Jones, L. R. *J. biol. Chem.* **266**, 11126–11130 (1991).
- Moorman, J. R., Palmer, C. J., John, J. E., Durieux, M. E. & Jones, L. R. *J. biol. Chem.* **267**, 14551–14554 (1992).
- Kowdley, G. C., Ackerman, S. J., John, J. E., Jones, L. R. & Moorman, J. R. *J. gen. Physiol.* **103**, 217–230 (1994).
- Attali, B. *et al. Nature* **365**, 850–852 (1993).
- Diamond, J. M. & Wright, E. M. *Physiol. Rev.* **49**, 581–646 (1969).
- Huxtable, R. J. *Physiol. Rev.* **72**, 101–163 (1992).
- Huxtable, R. & Chubb, J. *Science* **198**, 409–411 (1977).
- Atlas, M., Bahl, J., Roeske, W. & Bressler, R. *J. molec. cell. Cardiol.* **16**, 311–320 (1984).
- Banderall, U. & Roy, G. *Am. J. Physiol.* **263**, C1200–C1207 (1992).
- Kirk, K., Ellory, J. C. & Young, J. D. *J. biol. Chem.* **267**, 23475–23478 (1992).
- Goldstein, L. & Davis, E. M. *Am. J. Physiol.* **267**, R426–R431 (1994).
- Jackson, P. S. & Strange, K. *Am. J. Physiol.* **265**, C1489–C1500 (1993).
- Takumi, T., Ohkubo, H. & Nakanishi, S. *Science* **242**, 1042–1045 (1988).
- Folander, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 2975–2979 (1990).
- Hausdorff, S. F., Goldstein, S. A. N., Ruskin, E. E. & Miller, C. *Biochemistry* **30**, 3341–3346 (1991).
- Morrison, B. W. & Leder, P. *Oncogene* **9**, 3417–3426 (1994).
- Morrison, B. W., Moorman, J. R., Kowdley, G. C., Jones, L. R. & Leder, P. *J. biol. Chem.* **270**, 2176–2182 (1995).
- Attali, B., Latter, H., Rachamim, N. & Garty, H. *Proc. natn. Acad. Sci. U.S.A.* **92**, 6092–6096 (1995).
- Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, MA, 1992).
- Calá, S. E. & Jones, L. R. *Biophys. J.* **66**, A226(1994). (Abstract.)
- Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, NY, 1988).
- Pang, I.-H. & Sternweis, P. C. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7814–7818 (1989).
- Mounsey, J. P. *et al. J. clin. Invest.* **95**, 2379–2384 (1995).

ACKNOWLEDGEMENTS. We thank M. Entman for suggesting that we investigate taurine. This work was supported by the NIH and AHA.

† To whom correspondence should be addressed.

FIG. 1 Infection of T cells by FDC-trapped HIV-1. CD4⁺ T cells and autologous FDC-bearing HIV immune complexes trapped *in vitro* were cultured and infection monitored using PCR. Proviral HIV-1 gag (115 bp) DNA was detected where FDC-bearing HIV-1 were used as the only source of virus for infection of the T cells. Lysates of decreasing numbers of ACH-2 cells, a chronically infected T-cell clone containing one copy of proviral DNA per cell, were run in parallel for comparison. The intensity of the signal from T cells infected by FDC-bearing HIV immune complexes appeared between the intensity obtained from 10³ and 10⁴ ACH-2 cells. As a control, β -globin DNA was amplified simultaneously (bottom). T+SEE, T cells activated with staphylococcal enterotoxin E; FDC~HIV(IC), FDC incubated with HIV-1_{IIIIB} immune complexes; FDC, control FDC (same donor) not incubated with HIV.

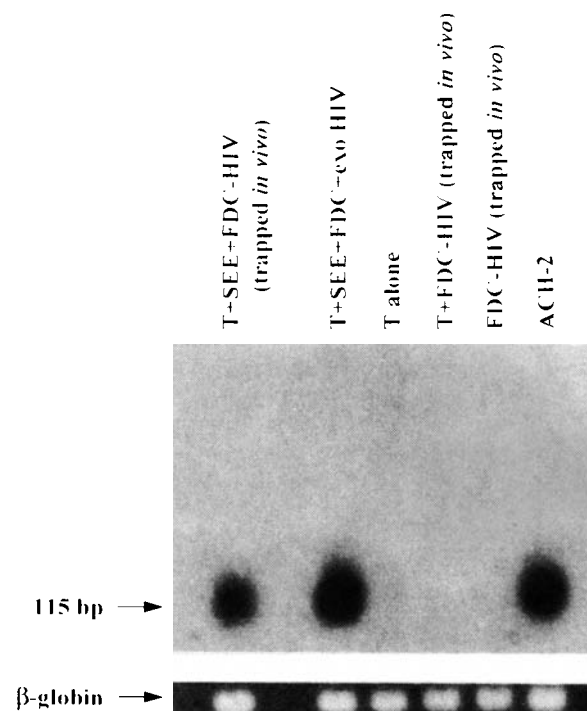
METHODS. Human FDC were obtained from tonsils of non-HIV-infected individuals using a modification of the procedure used for murine FDC^{9,13}. The modification was to increase incubation times and enzyme volumes to accommodate the larger mass of tissue present in the tonsils. The resulting FDC show the typical dendritic morphology, structure and function. FDC were further enriched using fluorescence-activated cell sorting (FACS) as described^{9,13}. FDC were stained using two (murine IgM, anti-human FDC) monoclonal antibodies: HJ2 (from M. Nahm^{14,15}) and DRC-1 (Dako). The FDC preparation was γ -irradiated (3,000 R) before incubation with HIV immune complexes to block proliferation and minimize the ability of the cells to support viral infection. FDC (6 $\times$ 10⁴) were incubated overnight (4 °C) with HIV immune complexes formed by incubating (for 2 h at 37 °C) 100- μ l fresh frozen serum from an HIV-infected individual (as a source of specific antibody and complement) with 100 μ l HIV_{IIIIB} cell-free supernatant (5,000 TCID₅₀). This dose of HIV provided sufficient virus for immune complex formation and subsequent trapping by FDC. HIV immune complexes not bound to FDC were removed by washing. FDC incubated with HIV immune complexes or control FDC were co-cultured in triplicate in 48-well plates, with FACS-purified, positively selected, autologous CD4⁺ T cells that had been activated overnight with SEE (100 μ g ml⁻¹; from Toxin Technology). After 4 days of culture, DNA was isolated as described¹⁶ and stored at -20 °C until tested. PCR analysis for HIV proviral gag and β -globin sequences was performed in the same reaction vessel using primer pairs SK38/39¹⁷ and GH20/PC04¹⁸, respectively. Control cultures of ACH-2 cells were isolated, diluted as indicated with uninfected H9 cells (these reagents were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH (ARRRP); ACH-2 was from



T. Folkes^{19,20}, H9 cells were from R. C. Gallo^{21,22}) to provide a constant amount of cellular DNA before isolation for PCR analysis. PCR amplification was performed for 35 cycles (94 °C for 2 min for first cycle, 92 °C for 1 min all other cycles; then 55 °C for 1.5 min; 72 °C for 2 min; after cycle 35, an additional 7-min incubation at 72 °C was performed). Amplified products (10 μ l) were analysed by electrophoresis (2% agarose), stained with ethidium bromide for β -globin DNA detection, and blotted onto Nytran. Blots were probed using ³²P-labelled SK19¹⁷. Autoradiography was for 3 d at -80 °C. Autoradiographs and Polaroid prints of ethidium bromide-stained gels were scanned at 400 dots per inch (d.p.i.), aligned using Microsoft Power Point, transferred to film and printed on Kodak F5 resin-coated paper.

FIG. 2 T-cell infection by HIV-1 trapped on murine FDC *in vivo*. Xenogeneic (murine) FDC bearing HIV immune complexes trapped *in vivo* were cultured with human CD4⁺ T cells and infection monitored using PCR. Infection of human T cells was clearly evident in cultures where FDC bearing *in vivo* trapped HIV were used as the only source of virus (T+SEE+FDC~HIV (trapped *in vivo*)). Infection was also detected when control murine FDC (no trapped HIV) were incubated with human T cells in the presence of exogenous HIV (5,000 TCID₅₀) and this signal appeared slightly more intense than that obtained from the FDC with *in vivo* trapped HIV. In contrast, no proviral DNA signal was obtained when the same FDC bearing virus were cultured with resting human T cells (T+FDC~HIV (trapped *in vivo*)), indicating the need for T-cell activation to support infection by the trapped virus. No proviral DNA signal was observed in control cultures of either the human T cells alone or the murine FDC-bearing HIV trapped *in vivo*. ACH-2 (5 $\times$ 10⁴) were included as a positive control. The ability of HIV on xenogeneic FDC to infect human T cells was not unexpected as there is no *a priori* reason why gp120 on FDC-trapped virus should not interact with CD4 present on human T cells. Furthermore, murine dendritic cells (DC, not to be confused with FDC) have also successfully infected human T cells *in vitro*¹⁶.

METHODS. HIV immune complexes were formed and trapped on FDC *in vivo* by injection (sc) of 1.2 mg per mouse, murine anti-gp120 mAb followed 2 h later by exposure to 600 R γ -irradiation to eliminate radiation-sensitive lymphocytes and thus enrich for radiation-resistant FDC^{9,23}. One day later, mice received injections of 5,000 TCID₅₀ HIV_{IIIIB} in the feet and behind the neck to distribute the virus to FDC in several lymph nodes. This dose of HIV resulted in consistent binding of HIV on FDC in draining lymph nodes. Five days later, when antigens including virus are confined to murine FDC²⁴, the mice were killed and the FDC isolated as described⁹. These preparations contain 25–45% FDC, with the remaining cells comprising equal numbers of T and B lymphocytes. Human peripheral blood CD4⁺ T cells were obtained from normal donors and enriched on lymphocyte-separation medium (Organon Technika), followed by sorting on the magnetic-activated cell sorter (MACS) using positive selection with anti-CD3 and anti-CD4 magnetic beads (Micro-



beads; Miltenyl Biotec, GmbH). Murine FDC (1 $\times$ 10⁵) bearing *in vivo* trapped HIV immune complexes (as the only source of virus) were co-cultured in triplicate with 1 $\times$ 10⁵ SEE-activated T cells for 4 days. PCR amplification was done as for Fig. 1.

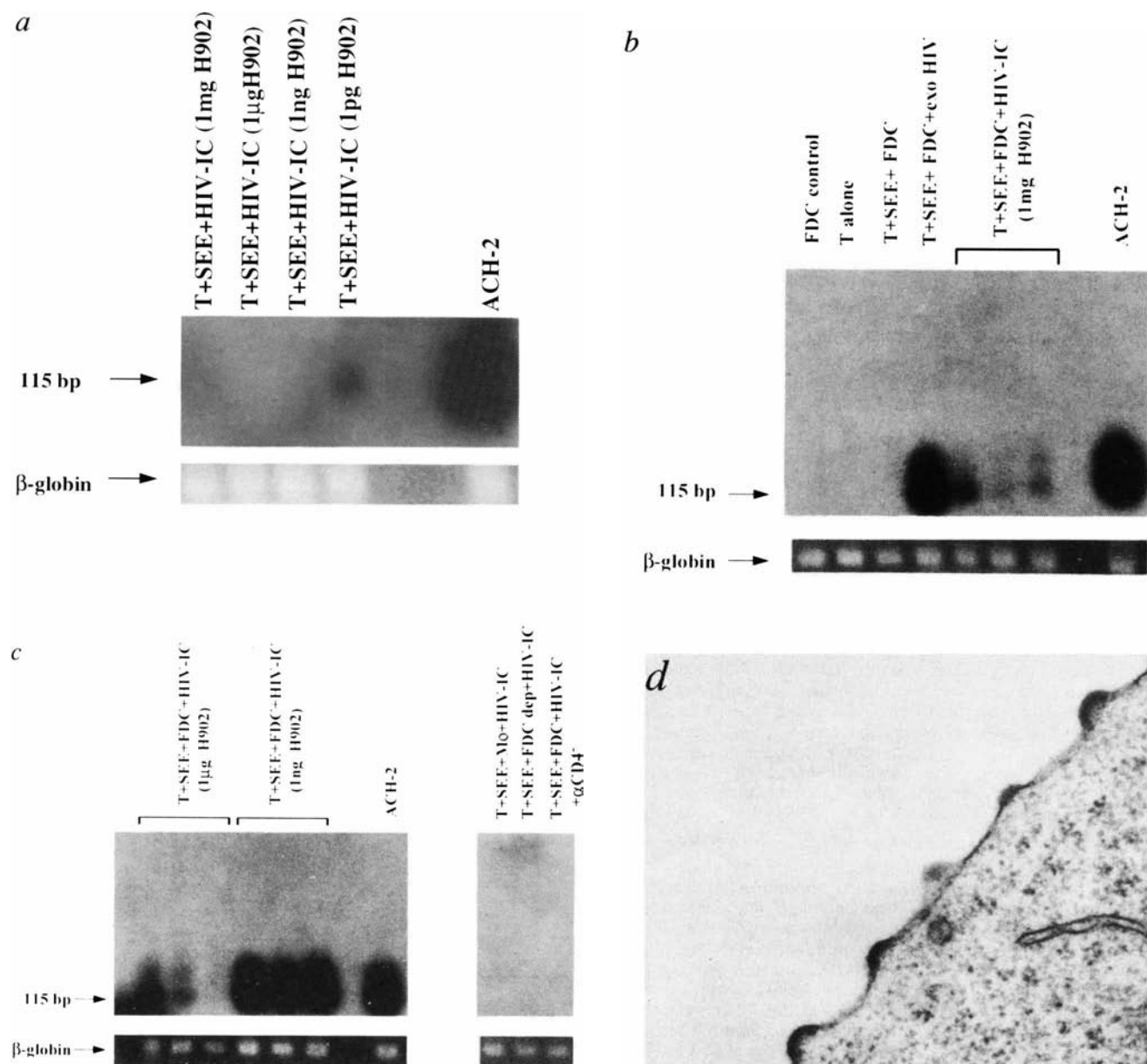


FIG. 3 FDC mediate HIV infection of T cells in the presence of neutralizing antibody. PCR amplification of HIV proviral DNA (top) and control β -globin DNA (bottom) from cultures of SEE-activated $CD4^+$ T cells (T+SEE) and HIV complexes (HIV-IC) formed with various doses of neutralizing antibody in the absence (a) or presence of human FDC (b and c). Infection was confirmed by electron microscopy (d). a, Infection of SEE-activated $CD4^+$ T cells by HIV-1_{III_B} immune complexes formed with 1 pg of neutralizing anti-gp120 (H902) was evident, whereas infection was blocked when viral complexes were formed using 1 ng, 1 μ g and 1 mg antibody. Signal intensity is compared with 50,000 ACH-2 cells. b and c (left), Addition of FDC to cultures reverses the effect of neutralizing antibody even in vast excess of the 1 ng needed to block infection when FDC were not present. c, right panel, Tonsillar macrophages (MO) and tonsillar cells depleted of FDC (FDC dep) could not substitute for FDC in promoting infection by HIV-IC formed with 1 ng of neutralizing antibody. Addition of anti-CD4 blocked infection even when FDC were present in cultures (+ $\alpha CD4$). d, Electron micrograph of a small portion of an infected cell from an FDC-T-cell culture containing HIV and 1 ng H902. Note the numerous HIV virions budding from the cell surface (magnification, $\times 54,000$). **METHODS.** All cells were isolated from human tonsils as described previously. SEE-activated T cells (5×10^4) were incubated at 37 °C for

1 h with HIV immune complexes formed using the indicated dose of neutralizing murine anti-HIV-1_{III_B} gp120 (the reagent, hybridoma 902 (anti-gp120) from B. Chesebro was obtained through the ARRRP^{10,25}) and 5,000 TCID₅₀ HIV_{III_B}. This dose of HIV was used to provide a high level of virus-anti-virus immune complexes, comparable to that which might exist on FDC *in vivo*⁴. All cultures were performed in triplicate and PCR analysis was as for Fig. 1. MACS was used to prepare tonsillar macrophages (positively selected using anti-CD14 Microbeads) and to remove FDC from tonsillar cells (using HJ2 followed by biotin-conjugated, rat anti-murine IgM and streptavidin Microbeads). Anti-CD4 (Leu 3a + 3b; Becton Dickinson) was added to cultures at a dose of 1 μ g. Cells for electron microscope evaluation were prepared using tannic acid during fixation²⁶.

these with HIV immune complexes *in vitro*. Viral immune complexes not bound to FDC were removed by washing and FDC were cultured with activated, autologous CD4⁺ T cells. Infection was monitored using the polymerase chain reaction (PCR) to analyse DNA from the cultures. Infection was demonstrated by isolation of proviral HIV-1 DNA from cultures of T cells and FDC-bearing HIV immune complexes, but not from cultures containing control FDC (Fig. 1). Syncytium formation was also observed in cultures of T cells and FDC-bearing HIV immune complexes. HIV binding to FDC was confirmed by electron microscopy, which also revealed the presence of infected cells budding virus (Fig. 3d). Portions of FDC dendritic processes bore 5–10 virus particles extracellularly, which is consistent with *in vivo* observations^{3,4,6}. The distribution of HIV virions on FDC processes suggests that a T cell can interact with several virus particles on a single portion of an FDC process.

Trapping of HIV immune complexes on FDC *in vitro* could differ from trapping *in vivo*, so we examined the infectivity of HIV trapped on FDC *in vivo* using a murine model. Previous work has shown that FDC can act as accessory cells to lymphocytes across species (G.F.B., unpublished) and we reasoned that the HIV-1 glycoprotein gp160 on FDC-trapped virus could interact with human CD4 on T cells regardless of FDC species. Antigens, including virus, can be localized on murine FDC *in vivo* by passive immunization followed by antigen challenge^{7,8}. FDC trapping of HIV was accomplished by passively immunizing mice with HIV-1_{IIIB}-specific anti-gp120, and one day later injecting 5,000 TCID₅₀ (half-maximal tissue culture infectious dose) HIV_{IIIB} in several sites to allow trapping of HIV immune complexes on FDC in multiple lymph nodes. Five days later, draining lymph nodes were obtained and the FDC isolated⁹. The murine FDC bearing HIV trapped *in vivo* were co-cultured with activated human T cells obtained from the blood of a normal donor (Fig. 2). Infection of human T cells was detected when HIV immune complexes on murine FDC were used as the only source of virus. Transfer of infection by murine FDC also indicated that HIV infection of FDC was not needed because the mouse is a non-permissive host. This model also excluded the possibility that human CD4⁺ cells in the FDC preparation were needed for transfer of infection.

The monoclonal antibody (H902) used to trap HIV on the murine FDC has been used to neutralize HIV-1_{IIIB} (ref. 10). The

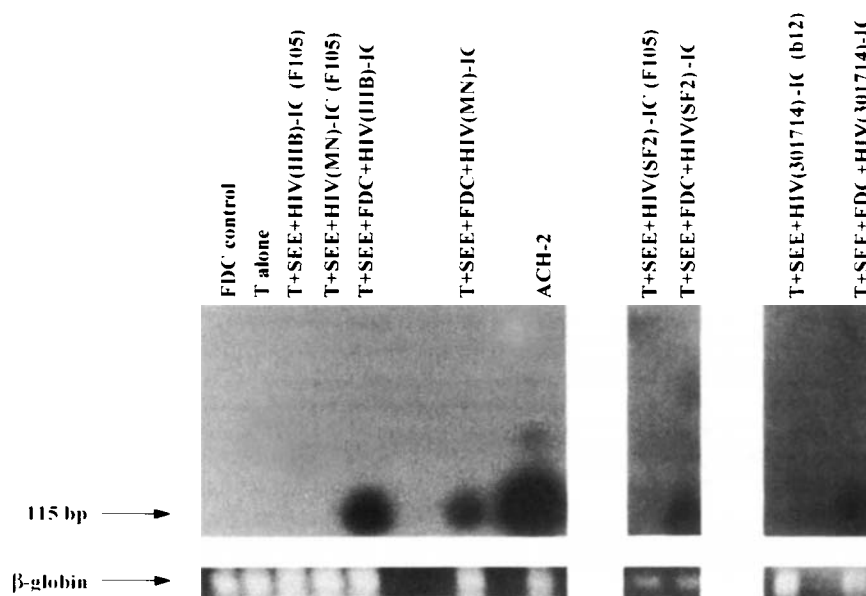
HIV complexes trapped on FDC *in vivo* were infectious: we therefore reasoned that FDC may be able to reverse the effect of neutralizing antibody. To test this, neutralized HIV-1_{IIIB} immune complexes were formed *in vitro* using increasing doses of neutralizing antibody and cultured with activated T cells in the presence or absence of FDC (Fig. 3). One nanogram of antibody consistently neutralized 5,000 TCID₅₀ HIV-1_{IIIB} (Fig. 3a). Infection occurred when FDC were present at doses of neutralizing antibody ranging from one thousand- to one million-fold above the one nanogram needed to prevent infection in the absence of FDC (Fig. 3b, c). Furthermore, infection was confirmed by electron microscopy which showed infected cells budding virus (Fig. 3d). Tonsillar macrophages and tonsillar cells specifically depleted of FDC failed to promote infection by neutralized HIV-1 immune complexes (Fig. 3c). Anti-CD4 in cultures of HIV immune complexes, T cells and FDC blocked infection, confirming the importance of the surface marker CD4. Increasing amounts of neutralizing antibody reduced the amount of proviral HIV-1 gag DNA detected, indicating that even though antibody did not block HIV on FDC, it did appear to decrease infection (Fig. 3b, c).

Immune complexes were formed with laboratory isolates HIV_{IIIB}, HIV-1_{MN} and HIV-1_{SF2}, the primary isolate 301714, and human neutralizing monoclonal antibodies (F105 or IgG1b12), in order to determine the general nature of infection by neutralized HIV on FDC (Fig. 4). In the absence of FDC, HIV-1 immune complexes were unable to infect T cells regardless of the strain of HIV-1 used. But, when FDC were present, neutralized virus was able to infect the T cells, indicating that this FDC-mediated process was not restricted to a single antibody or HIV-1 strain. It is noteworthy that the laboratory isolate SF2, which is easily neutralized by antibody, is rendered infectious when FDC are present. Thus under conditions where antibody is particularly efficient at neutralizing virus, FDC can still negate its effect—suggesting that this FDC activity is very potent.

The ability of FDC to facilitate HIV infection in the presence of neutralizing antibody may help to explain why some individuals with a high titre of neutralizing antibody have ongoing infection. It may also help to explain why much of the viral replication occurring during the clinically latent stage of HIV infection is confined to lymphoid follicles where virus-laden FDC reside in intimate contact with germinal centre T and B

FIG. 4 FDC-mediated infection by neutralized HIV-1 immune complexes is not restricted to viral strain or antibody. PCR analysis of HIV-1 gag proviral DNA (115 bp) and β -globin in cultures containing various strains of neutralized HIV-1 immune complexes (HIV-IC) and SEE-activated T cells (T + SEE) (2.6×10^4) or SEE-activated T cells and human FDC (5×10^4). Infection by HIV-1 immune complexes formed with human, neutralizing anti-gp120 (1 μ g F105 or 20 μ g IgG1 b12) was blocked in the absence of FDC with HIV-1 strains IIIB, MN and SF2 as reported (ARRRP data sheet) and with the primary strain 301714. In contrast, infection was apparent in cultures when FDC were present, indicating that the FDCs ability to reverse the effect of neutralizing antibody was not restricted to a single strain of virus or monoclonal antibody.

METHODS. Cultures of SEE-activated T cells and MACS-sorted FDC were incubated with HIV-1 immune complexes and assessed for infection as described for Fig. 3, with the exception that the HIV immune complexes were formed using HIV-1 strains (IIIB, MN and SF-2) (obtained through the ARRRP: strains IIIB and MN from R. Gallo^{22,26,27}; strain SF2 from J. Levy²⁸) mixed with 1 μ g human anti-HIV-1 gp120 (obtained through the ARRRP: HIV-1 gp120 monoclonal antibody (F105) from M. Posner^{27,28}). Immune complexes were also prepared using primary isolate 301714 (obtained from



the ARRRP) and monoclonal antibody b12 (ref. 29; from D. R. Burton and P. Parren).

cells. Although the mechanism for converting neutralized virus into an infectious form by FDC is unclear, there are similarities to events that occur in anamnestic responses, in which antigens are trapped by FDC in the form of immune complexes. Even though antibody should 'mask' epitopes and prevent recognition by the immunoglobulin receptors of B cells, experiments have shown that antigen-specific B cells are able to recognize the immune-complexed antigen on FDC or iCCosomes even in vast antibody excess^{11,12}. We envisage a similar process in the case of FDC-trapped HIV. The V3 loop of HIV-1 gp160 would potentially be 'masked' by neutralizing antibody, but the FDC can display the virus in such a manner that gp120 can bind CD4 on adjacent T cells and thus cause infection. These possible mechanisms for negating the effect of neutralizing antibody are being actively pursued. Nevertheless, the finding that FDC can convert neutralized HIV immune complexes into an infectious form may have important implications for the design of therapeutic and vaccine strategies, regardless of the mechanism(s) involved. □

Received 24 July; accepted 7 September 1995.

1. Racz, P., Tenner-Racz, K. & Schmidt, H. *Acta path. microbiol. immun. Scand.* (suppl.) **8**, 16–23 (1989).
2. Biberfeld, P. et al. *Cancer Detect. Prev.* **12**, 217–224 (1988).
3. Pantaleo, G. et al. *Nature* **362**, 355–358 (1993).

4. Embretson, J. et al. *Nature* **362**, 359–362 (1993).
5. Fox, C. H. & Cottler-Fox, M. *Immun. Today* **13**, 353–356 (1992).
6. Tenner-Racz, K. et al. in *Accessory Cells in HIV and Other Retroviral Infections* (eds Racz, P., Dijkstra, C. D. & Gluckman, J. C.) Vol. 1, 83–97 (Karger, Basel, 1991).
7. Tew, J. G., Mandel, T. E. & Miller, G. A. *Aust. J. exp. Biol. med. Sci.* **57**, 401–414 (1979).
8. Szakal, A. K., Holmes, K. L. & Tew, J. G. *J. Immun.* **131**, 1714–1727 (1983).
9. Burton, G. F., Conrad, D. H., Szakal, A. K. & Tew, J. G. *J. Immun.* **150**, 31–38 (1993).
10. Chesebro, B. & Wehrly, K. J. *Virology* **62**, 3779–3788 (1988).
11. Gray, D., Kosco, M. H. & Stockinger, B. *Int. Immun.* **3**, 141–148 (1991).
12. Tew, J. G., Greene, E. J. & Makoski, M. H. *Cell Immun.* **26**, 141–152 (1976).
13. Tew, J. G., Burton, G. F. & Szakal, A. K. in *Handbook of Experimental Immunology: The Lymphoid System Antibody Responses and Affinity Maturation* (ed. Nossal, G. J. V.) (Rockwell Scientific, Oxford, in the press).
14. Butch, A., Chung, G., Hoffmann, J. & Nahm, M. J. *Immun.* **150**, 39–47 (1993).
15. Butch, A. W., Hug, B. A. & Nahm, M. H. *Cell Immun.* **155**, 27–41 (1994).
16. Camron, P. U. et al. *Science* **257**, 383–387 (1992).
17. Ou, C. et al. *Science* **239**, 295–297 (1988).
18. Bauer, H. M. et al. *J. Am. med. Soc.* **265**, 472–477 (1991).
19. Clouse, K. A. et al. *J. Immun.* **142**, 431–438 (1989).
20. Folks, T. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2365–2368 (1989).
21. Mann, D. L. et al. *AIDS Res. hum. Retrovir.* **5**, 253–255 (1989).
22. Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. *Science* **224**, 497–500 (1984).
23. Kosco, M. H., Pflugfelder, E. & Gray, D. J. *Immun.* **148**, 2331–2339 (1992).
24. Mandel, T. E., Phipps, R. P., Abbot, A. & Tew, J. G. *Immun. Rev.* **53**, 29–59 (1980).
25. Pincus, S., Wehrly, K. & Chesebro, B. J. *Immun.* **142**, 3070–3075 (1989).
26. Gelderblom, H. R., Hausmann, E. S. H., Ozel, M., Pauli, G. & Koch, M. A. *Virology* **156**, 171–176 (1987).
27. Posner, M. R., Cavacini, L. A., Emes, C. L., Power, J. & Byrn, R. J. *Acqu. Immun. Defic. Syndr.* **6**, 7–14 (1995).
28. Posner, M. R. et al. *J. Immun.* **146**, 4325–4332 (1991).
29. Burton, D. R. et al. *Science* **266**, 1024–1027 (1994).

ACKNOWLEDGEMENTS. We thank C. Bass for technical assistance, V. Bruzzese and T. Kerker for patient sera, and G. Bradley for helpful suggestions and criticism. This work was supported by grants from the NIAID.

Behavioural and cardiovascular effects of disrupting the angiotensin II type-2 receptor gene in mice

Lutz Hein*, Gregory S. Barsh†‡, Richard E. Pratt*, Victor J. Dzau* & Brian K. Kobilka*†§

* Falk Cardiovascular Research Center and Department of Medicine, † Departments of Genetics and Pediatrics, and ‡ Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA

ANGIOTENSIN II, a potent regulator of blood pressure and of water and electrolyte balance, binds to two different G-protein-coupled receptors. The type-1 receptor (AT₁) mediates the vasopressor and aldosterone-secreting effects of angiotensin II, but the function of the type-2 receptor (AT₂; refs 1, 2) is unknown, although it is expressed in both adult³ and embryonic⁴ life. To address this question, we have generated mice lacking the gene encoding the AT₂ receptor. Mutant mice develop normally, but have an impaired drinking response to water deprivation as well as a reduction in spontaneous movements. Their baseline blood pressure is normal, but they show an increased vasopressor response to injection of angiotensin II. Thus, although the AT₂ receptor is not required for embryonic development, it plays a role in the central nervous system and cardiovascular functions that are mediated by the renin-angiotensin system.

The angiotensin II AT₂ receptor gene (*Agtr2*) on the X chromosome^{5,6} was disrupted in embryonic stem (ES) cells to generate mice deficient in this receptor subtype (Fig. 1). Crossing heterozygous female mice with a male germline-transmitting chimera gave 24/94 non-mutant (+/Y), 21/94 heterozygous (-/+), 28/94 homozygous (-/-) and 21/94 hemizygous (-/Y) mutant mice. This distribution does not deviate significantly from mendelian prediction (1/4:1/4:1/4:1/4; $\chi^2 = 1.405$,

$P > 0.05$), indicating that there was no increase in embryonic or perinatal mortality in mice of the *Agtr2*^{-/-} or *Agtr2*^{-/Y} genotype. Mice carrying the mutated *Agtr2* allele apparently developed normally. Histological sections of heart, lung, kidney, adrenal, spleen, brain, aorta, ovary, uterus, pancreas, eye, skeletal muscle and blood did not reveal any differences in morphology between non-mutant and hemizygous or homozygous mutant mice (data not shown). The skeletal system was differentially stained with alizarin red S for bone and alcian blue for cartilage in 1-day-old *Agtr2* mutant mice and showed no sign of malformation or retarded ossification (data not shown). Furthermore, homozygous mutant females and hemizygous mutant males (age 8–16 weeks) produced healthy litters of normal size (average size of litter, 6.1 ± 0.6 (*Agtr2*^{-/-} × *Agtr2*^{-/Y}; $n = 7$) versus 6.0 ± 0.5 (*Agtr2*^{-/+} × *Agtr2*^{-/Y}; $n = 15$). $P > 0.05$).

To demonstrate that functional AT₂ receptors were missing in *Agtr2* mutant mice, their RNA was analysed and radioligand binding experiments were done on membranes prepared from whole mouse fetuses (minus placenta) on day 18.5 of embryonic development (E18.5). In normal rat embryos, the level of AT₂ receptor expression is maximal a few days before birth^{4,7}. We detected no intact AT₂ messenger RNA in hemizygous or homozygous *Agtr2* mutant E18.5 mice (Fig. 1c). Binding studies confirmed the expression of AT₂ receptors in non-mutant E18.5 mice: males had 725 ± 21 fmol mg⁻¹ ($n = 6$) and females 703 ± 29 fmol mg⁻¹ (Fig. 1d). In heterozygous *Agtr2*^{-/+} female embryos, AT₂ receptor expression was 395 ± 38 fmol mg⁻¹ ($n = 8$), approximately half that in non-mutant mice; in homozygous female or hemizygous male fetuses, none was detected (Fig. 1d). Expression of AT₁ receptors did not differ between non-mutant mice and mice carrying the mutated *Agtr2* allele (data not shown).

Angiotensin II influences several central nervous system functions, including dipsogenesis, activation of the sympathetic nervous system, modulation of hormone release from the pituitary, interference with learning and motor activity, and regulation of body temperature and analgesia⁸. Although AT₁ and AT₂ receptors may both participate in the dipsogenic response to angiotensin II (refs 9–12), the function of AT₂ is unclear^{10,13}. We therefore examined the drinking behaviour of knockout mice and found no difference in daily water intake compared with

§ To Whom correspondence should be addressed.

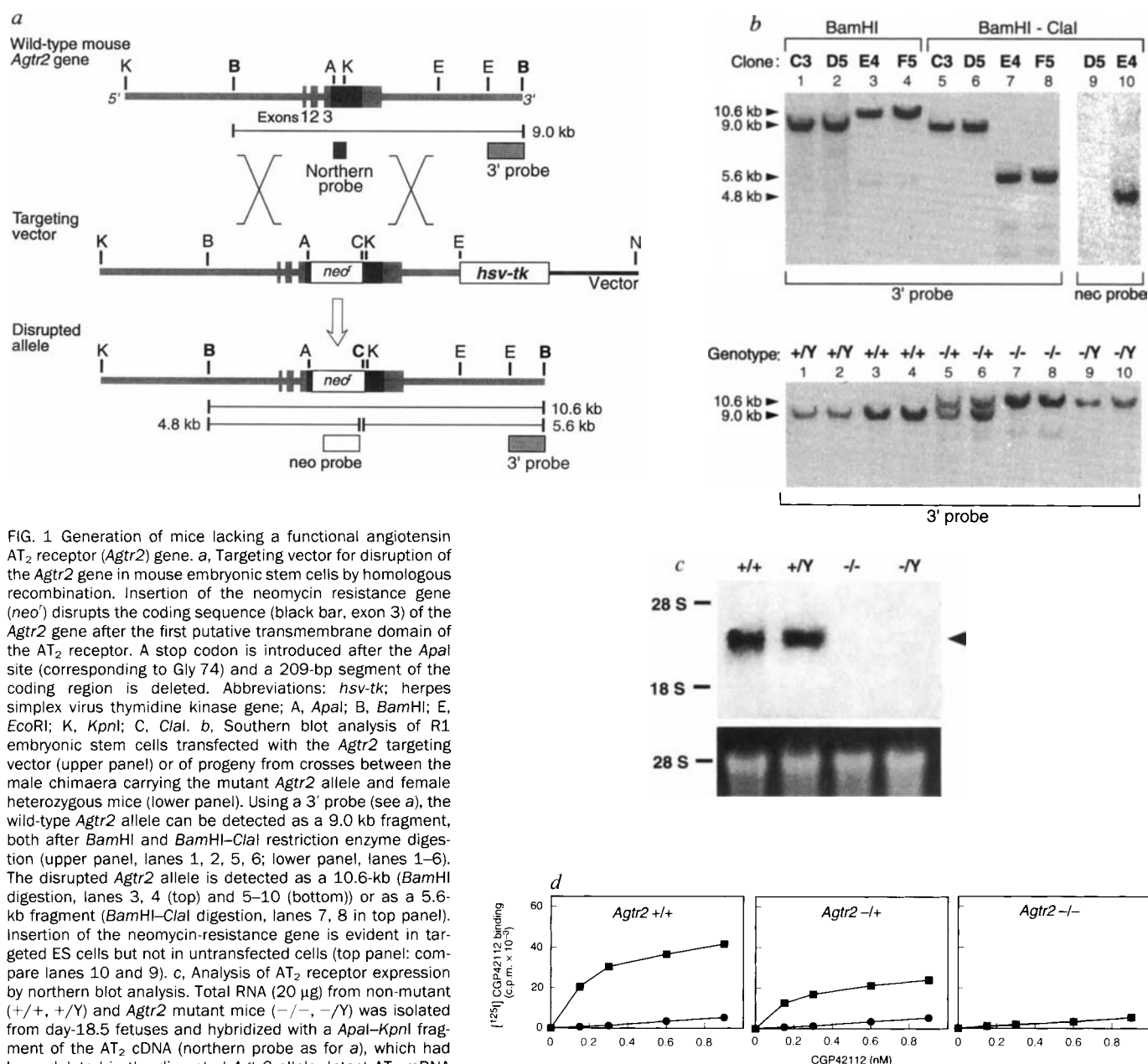


FIG. 1 Generation of mice lacking a functional angiotensin AT₂ receptor (*Agtr2*) gene. **a**, Targeting vector for disruption of the *Agtr2* gene in mouse embryonic stem cells by homologous recombination. Insertion of the neomycin resistance gene (*neo*) disrupts the coding sequence (black bar, exon 3) of the *Agtr2* gene after the first putative transmembrane domain of the AT₂ receptor. A stop codon is introduced after the *Apal* site (corresponding to Gly 74) and a 209-bp segment of the coding region is deleted. Abbreviations: *hsv-tk*; herpes simplex virus thymidine kinase gene; A, *Apal*; B, *Bam*HI; E, *Eco*RI; K, *Kpn*I; C, *Clal*. **b**, Southern blot analysis of R1 embryonic stem cells transfected with the *Agtr2* targeting vector (upper panel) or of progeny from crosses between the male chimaera carrying the mutant *Agtr2* allele and female heterozygous mice (lower panel). Using a 3' probe (see **a**), the wild-type *Agtr2* allele can be detected as a 9.0 kb fragment, both after *Bam*HI and *Bam*HI-*Clal* restriction enzyme digestion (upper panel, lanes 1, 2, 5, 6; lower panel, lanes 1–6). The disrupted *Agtr2* allele is detected as a 10.6-kb (*Bam*HI digestion, lanes 3, 4 (top) and 5–10 (bottom)) or as a 5.6-kb fragment (*Bam*HI-*Clal* digestion, lanes 7, 8 in top panel). Insertion of the neomycin-resistance gene is evident in targeted ES cells but not in untransfected cells (top panel: compare lanes 10 and 9). **c**, Analysis of AT₂ receptor expression by northern blot analysis. Total RNA (20 µg) from non-mutant (+/+), (+/Y) and *Agtr2* mutant mice (–/–, –/Y) was isolated from day-18.5 fetuses and hybridized with a *Apal*-*Kpn*I fragment of the AT₂ cDNA (northern probe as for **a**), which had been deleted in the disrupted *Agtr2* allele. Intact AT₂ mRNA (arrowhead) is absent in *Agtr2* mutant mice. Lower panel, gel stained with ethidium bromide before transfer to show that similar amounts of RNA were analysed from mutant and non-mutant mice. Migration positions of ribosomal RNA (18S and 28S) are shown. **d**, Membranes from whole E18.5 fetuses (minus placenta) were prepared and AT₂ receptors detected by binding of the radioligand [¹²⁵I]-CGP42112. Results from saturation binding experiments are shown as total [¹²⁵I]-CGP42112 binding (squares) and nonspecific binding in the presence of 3 µM Sar-Ile-angiotensin II (circles). Representative data are shown for non-mutant mice, heterozygous females and homozygous female mutants. For homozygous females, the plots for total binding and nonspecific binding are identical, demonstrating the absence of functional AT₂ receptors in these mice.

METHODS. The genomic DNA for the mouse AT₂ receptor (*Agtr2*) was cloned from a 129/Sv mouse genomic library (Stratagene) by conventional plaque hybridization using mouse cDNA as a probe²⁶. For the gene targeting vector, a 6.6-kb genomic DNA fragment (*Kpn*I-*Apal*) containing the promoter, complete exons 1 and 2 and the beginning of the third exon, was placed upstream of a neomycin-resistance gene²⁷ in a pBluescript SK-plasmid. For the 3' part of the targeting vector, a 2.3-kb *Kpn*I-*Eco*RI fragment, which included the AT₂ coding sequence after the second transmembrane segment and the 3' untranslated region, was linked to the HSV thymidine kinase gene²⁷. For the targeting experiments, R1 embryonic stem cells (derived from 129/Sv mice)²⁸ were electroporated with 20 µg targeting plasmid previously linearized by digestion with *Not*I. R1 cells were cultured at 37 °C in 95% air/5% CO₂ on confluent monolayers of irradiated primary embryonal fibroblasts in Dulbecco's modified Eagle's medium (DME-H21; UCSF tissue

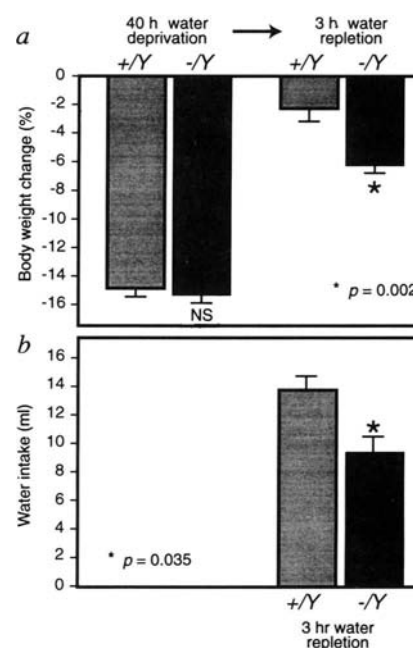
culture facility, San Francisco) supplemented with 20% fetal bovine serum (Hyclone), 1 mM sodium pyruvate (Gibco BRL), non-essential amino acids and penicillin/streptomycin (1x; UCSF Cell Culture Facility), 10^{–4} M β-mercaptoethanol and 2,000 U ml^{–1} leukaemia inhibitory factor (ESGRO; Gibco BRL). G418- and gancyclovir-resistant ES cell clones were analysed by Southern blotting. Following hybridization with the ³²P-labelled 3' *Agtr2* probe, homologous recombination events were identified by a shift in band size from 9.0 kb (wild-type allele) to 10.6 kb (targeted allele) after *Bam*HI digestion. Chimaeric mice were produced by aggregation of morulae with targeted ES cells as described²⁹. Briefly, groups of 10–12 targeted ES cells were incubated overnight in single droplet culture with morulae collected from pregnant CD-1 female mice on day 2.5 after plugging. Blastocysts (20–25) were transferred into the left uterine horn of pseudopregnant female CD-1 mice. To obtain germline transmission of the targeted *Agtr2* allele, male chimaeric mice were mated with female wild-type FVB/N mice. This cross produced female mice heterozygous for the *Agtr2* disruption and these heterozygous females were subsequently mated to male FVB/N wild-type mice to produce the F₂ generation. Experiments shown in Figs 2–4 were carried out on F₂ generation male mutant (*Agtr2*–/Y) or non-mutant (*Agtr2*+ /Y) mice. Expression of the AT₂ receptor was tested in wild-type mice and knockout mice by radioligand binding as described². For this purpose, membranes were prepared from E18.5 mice. For detection of AT₂ receptors, the ligand [¹²⁵I]-CGP42112 was used in saturation binding experiments. Nonspecific binding was determined in the presence of 3 µM Sar-Ile-angiotensin II. Experiments were repeated in five litters with identical results.

FIG. 2 Body weight and drinking response of *Agtr2* mutant mice following water deprivation. Change in body weight (as a percentage of initial weight; a) and water intake (as ml per 100 g body weight; b) were recorded after a 40-hour period of water deprivation. During this time the weight loss of the *Agtr2* mutant mice and the non-mutant normal mice was similar (a). Three hours after the return of drinking water, non-mutant mice had regained significantly more of their body weight than the *Agtr2* mutant mice; water intake was higher in normal mice than in knockout mice (b).

METHODS. For a, 18 mutant and 18 non-mutant male mice were studied. The mice were from 13 different litters (Fig. 1 legend), each litter contributing 1 or 2 mutant and 1 or 2 non-mutant mice. During the period of water deprivation, mice had free access to food. Mice were weighed before and at the end of this period, and 3 h after the return of drinking water. In b, the drinking response to water deprivation was measured in 6 groups of mutant mice and 6 groups of non-mutant mice. Each group consisted of 2 or 3 male mice from 10 different litters, each litter contributing 1 or 2 mutant and 1 or 2 non-mutant mice (a total of 14 *Agtr2*^{-/-} and 14 *Agtr2*^{+/+}). After the water deprivation period, water was offered in 50-ml polypropylene tubes with metal drinking sippers attached. Water intake was measured by weighing the water tubes before and after the experiment. For experiments shown in a and b, mutant and non-mutant mice were kept in separate cages to avoid competition for water or food. Values are means $\pm$ s.e.m. for $n=18$ mice (a) and $n=6$ groups of mice (b); the Mann-Whitney test was used for statistical comparison between groups.

normal mice when water was freely available (data not shown). But after 40 hours without water, the drinking response during the ensuing 3 hours was significantly impaired in the *Agtr2* mutant mice compared to normals ($P=0.035$; Fig. 2); changes in body weight were consistent with these observations. The loss in body weight due to water deprivation was identical in knockout and non-mutant mice (*Agtr2*^{+/+}: $-15.3 \pm 0.6\%$; *Agtr2*^{-/-}: $-14.9 \pm 0.6\%$; $n=18$), but restoration of body weight was significantly reduced ($P=0.002$) in the mutant mice (Fig. 2a), suggesting that the AT₂ receptor subtype is important in regulating drinking behaviour after water deprivation.

Intracerebroventricular injection of angiotensin II stimulates exploratory behaviour and locomotor activity^{14,16}. In rats the



AT₂ subtype is strongly expressed in the locus coeruleus^{17,18}, and stimulation of the locus coeruleus by angiotensin II causes release of noradrenaline in brain slices^{19,20}. As the locus coeruleus is involved in the integration of sensory information and in arousal, we tested whether the absence of functional AT₂ receptors in knockout mice affected their locomotor activity. Measurement of locomotion in a photobeam cage system revealed a significant decrease in the activity of the *Agtr2* mutant mice compared with non-mutant mice (Fig. 3), particularly during the first 6 hours in a new environment (day 1 in Fig. 3a). There was a similar but smaller effect during the dark period of the light/dark cycle on the second day in the activity cage. The physiological significance of the AT₂-receptor-mediated locomotor activity is

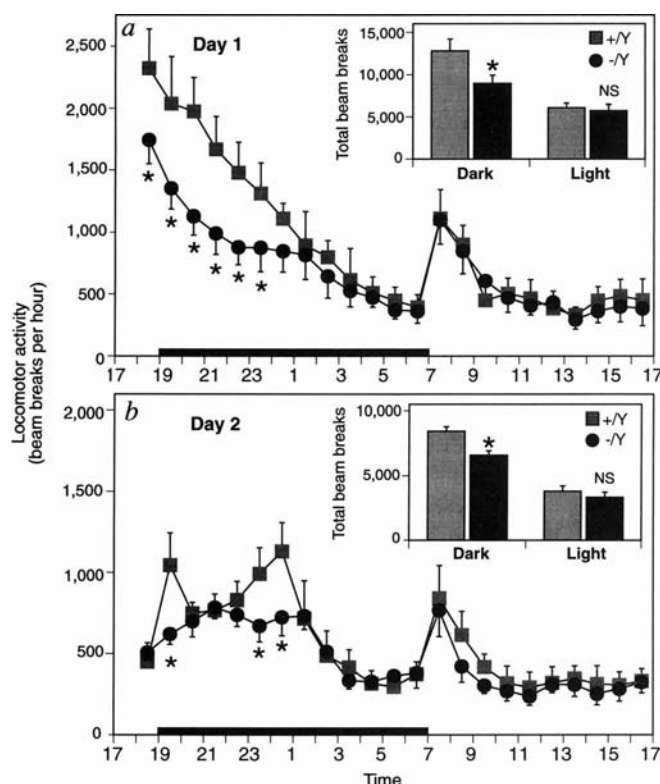


FIG. 3 Locomotor activity of *Agtr2* mutant mice. Pairs of male littermates (non-mutant *Agtr2*^{+/+}, mutant *Agtr2*^{-/-}) were monitored in a photobeam activity cage for two consecutive days. During the dark period of the day/night cycle, *Agtr2* mutant mice were significantly less active (inserts in a, b), although there was no difference between the groups during the light period. The reduced locomotor activity of the *Agtr2* mutant mice during the dark period is apparent during the first 6 hours after the start of the test (a, day 1) as well as in the first half of the dark period on the second test day (b, day 2).

METHODS. Locomotor activity was followed using a photobeam cage system (San Diego Instruments) as described³⁰. Two standard polypropylene cages (30 $\times$ 50 cm) were placed inside adjustable frames equipped with 4 $\times$ 6 infrared photobeams. For each cage, all beam breaks were relayed by an interface to a PC386-based microcomputer. Pairs of mutant and non-mutant male littermates were kept in the same cage from the time of weaning (age 3 weeks) until the time of the experiments (age 11–14 weeks). No other experiments were done on these animals before the behavioural studies. This protocol eliminates any possible habituation of the mice to different environments beforehand. Matched pairs of male littermate mice were placed in the photobeam cages for a 24-hour period, starting at 17:00. Data were recorded as total number of beam breaks in 5-min intervals. Locomotor activity was analysed by two-way analysis of variance (ANOVA) followed by the Mann-Whitney test for statistical significance between groups ($n=10$ pairs of littermates; NS, not significant; *, significant at $P=0.02$).

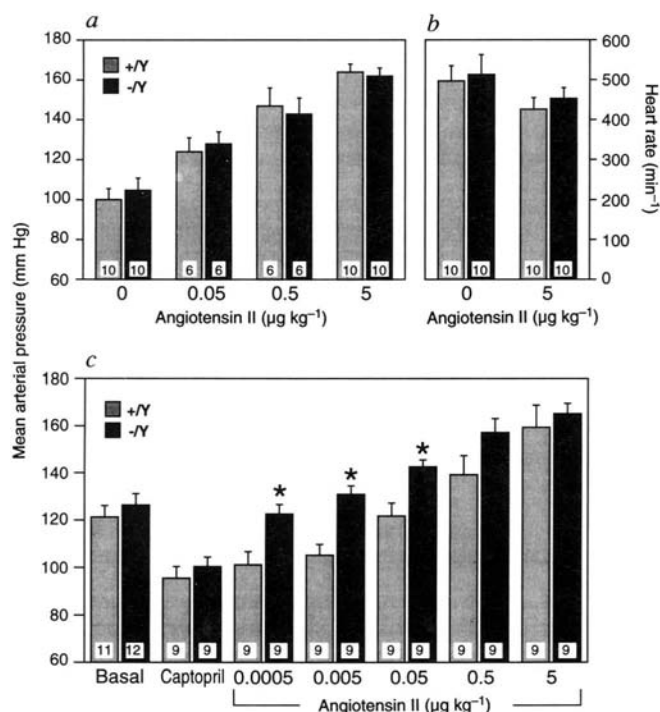


FIG. 4 Blood pressure (a, c) and heart rate (b) of *Agtr2* mutant mice. Blood pressure was measured in conscious, unrestrained mice 18–24 h after catheterization of the left carotid artery. Basal mean aortic pressure did not differ significantly between non-mutant male mice and hemizygous *Agtr2*^{-/-} mice at 13–14 weeks of age (a; $P=0.48$, unpaired *t*-test) or at 28–30 weeks of age (c; $P=0.69$, unpaired *t*-test). Injection of angiotensin II resulted in similar dose-dependent increases in mean aortic pressures in both groups (a; $P=0.63$, two-way analysis of variance ANOVA). After injection of captopril (30 mg kg^{-1}) to inhibit endogenous production of angiotensin II, the mean arterial pressure decreased to similar levels in both groups (c). Under these conditions, the pressure responses to angiotensin II given at lower doses (0.0005 – 0.05 µg kg^{-1}) were enhanced in *Agtr2*^{-/-} mice compared with *Agtr2*^{+/+} mice (c; $P=0.028$, two-way analysis of variance ANOVA followed by Bonferroni tests, $*P<0.02$). Boxed numbers at the base of each column indicate the number of animals tested.

METHODS. Blood pressure was measured in chronically instrumented, conscious and unrestrained mice. For these experiments, male mice (a, aged 13–14 weeks; b, 28–30 weeks) were anaesthetized with methoxyflurane and a catheter was inserted into the left common carotid artery. After overnight recovery, the catheter was connected to a Statham pressure transducer element. Angiotensin II was applied at different concentrations (1 nM – 10 µM) in a 10-µl volume directly into the catheter. Data were registered simultaneously on a Gould chart recorder and transmitted on-line via a Gould amplifier to a 486PC-computer.

unclear, but the impaired drinking response in knockout mice may be due not only to the effect of the disrupted gene on drinking behaviour but also to a decrease in their general motor activity.

To test whether the AT_2 receptor plays a role in blood pressure regulation, arterial blood pressure was measured in unrestrained and conscious normal and knockout mice using an arterial catheter in the left carotid artery. Basal aortic blood pressure and heart rate did not differ significantly between non-mutant and hemizygous male knockout mice (Fig. 4a, b). Likewise, injection of pharmacological doses of angiotensin II resulted in a similar dose-dependent increase in arterial blood pressure in both groups, suggesting that the pressor response mediated by the AT_1 receptor was not affected by disruption of the *Agtr2* gene (Fig. 4a). These results differ from those of Ichiki *et al.*²¹, who use a different protocol for angiotensin II administration. They pretreated mice with captopril to block production of endogenous angiotensin II, then continuously infused lower and more physiological doses of angiotensin II. Using a similar protocol, we also find an increased sensitivity to low-dose angiotensin II in knockout animals (Fig. 4c), but we cannot detect any significant difference in baseline blood pressure compared with control littermates (Fig. 4a). This discrepancy in the effect of the AT_2 receptor knockout on baseline blood pressure could be due to differences in the genetic background of the mouse strains, or, as wild-type and mutant mice differ in their sensitivity to angiotensin II, factors such as diet, environmental stress or age might affect baseline blood pressure in our animals.

Inactivation of the angiotensin AT_2 receptor gene *Agtr2* in mice by gene targeting indicates that this receptor is not required for normal development but is responsible for behavioural and cardiovascular effects of angiotensin II in adult mice. Other genes closely linked to the disrupted AT_2 locus on the X chromosome may also be involved in these mice, which are produced from a cross of 129/Sv and FVB/N strains, although no other candidate genes for the observed phenotypes have been identified around the locus of the AT_2 receptor gene⁵.

The role of the AT_2 receptor in selective functions of the central nervous system and in cardiovascular regulation may have pharmacotherapeutic implications. In addition to the sites of

tissue expression in the normal adult, AT_2 receptors are re-expressed pathologically in neointimal hyperplasia after vascular injury²², in myocardial infarction and cardiac hypertrophy^{23,24}, and in skin wounds²⁵. This animal model may be useful for investigation of the role of the AT_2 receptor subtype in these pathophysiological conditions and for testing therapies for these conditions. □

Received 6 June; accepted 14 September 1995.

- Kambayashi, Y. *et al.* *J. Biol. Chem.* **268**, 24543–24546 (1993).
- Mukoyama, M. *et al.* *J. Biol. Chem.* **268**, 24539–24542 (1993).
- Zhuo, *et al.* in *Hypertension: Pathophysiology, Diagnosis, and Management* (eds Laragh, J. H. & Brenner, B. M.) 1739–1762 (Raven, New York, 1995).
- Grady, E. F., Sechi, L. A., Griffin, C. A., Schambelan, M. & Kalinyak, J. E. *J. Clin. Invest.* **88**, 921–933 (1991).
- Chassagne, C., Beatty, B. G. & Meloche, S. *Genomics* **25**, 601–603 (1995).
- Koike, G. *et al.* *Biochem. biophys. Res. Commun.* **203**, 1842–1850 (1994).
- Feuillan, P. P., Millan, M. A. & Aguilera, G. *Regul. Pept.* **44**, 159–169 (1993).
- Wright, J. W. & Harding, J. W. *Neurosci. Biobehav. Rev.* **18**, 21–53 (1994).
- Cooney, A. S. & Fitzsimons, J. T. *Expl. Physiol.* **78**, 767–774 (1993).
- Hogarty, D. C., Speakman, E. A., Puig, V. & Phillips, M. I. *Brain Res.* **586**, 289–294 (1992).
- Rowland, N. E., Rozelle, A., Riley, P. J. & Fregly, M. J. *Brain Res. Bull.* **29**, 389–393 (1992).
- Rowland, N. E. & Fregly, M. J. *Brain Res. Bull.* **32**, 391–394 (1993).
- Dourish, C. T., Duggan, J. A. & Banks, R. J. *Eur. J. Pharmac.* **211**, 113–116 (1992).
- Georgiev, V. *Pol. J. Pharmac. Pharm.* **42**, 553–562 (1990).
- Georgiev, V., Getova, D. & Opitz, M. *Meth. Find. Exp. Clin. Pharmac.* **13**, 5–9 (1991).
- Petkova, B., Georgiev, V. & Kambourova, T. *Acta Physiol. Pharmac. Bulg.* **17**, 84–90 (1991).
- Rowe, B. P., Saylor, D. L. & Speth, R. C. *Neuroendocrinology* **55**, 563–573 (1992).
- Tsutsumi, K. & Saavedra, J. M. *Molec. Pharmac.* **41**, 290–297 (1992).
- Sumners, C. & Phillips, M. I. *Am. J. Physiol.* **244**, R257–R263 (1983).
- Huang, Y., Rogers, J. & Henderson, G. J. *Neurochem.* **49**, 1541–1549 (1987).
- Ichiki, T. *et al.* *Nature* **377**, 748–750 (1995).
- Viswanathan, M., Stromberg, C., Seltzer, A. & Saavedra, J. M. *J. Clin. Invest.* **90**, 1707–1712 (1992).
- Lopez, J. J. *et al.* *Am. J. Physiol.* **H844**–**H852** (1994).
- Nio, Y., Matsubara, H., Murasawa, S., Kanasaki, M. & Inada, M. *J. Clin. Invest.* **95**, 46–54 (1995).
- Grady, E. F. & Kalinyak, J. E. *Regul. Pept.* **44**, 171–180 (1993).
- Nakajima, M., Mukoyama, M., Pratt, R. E., Horiuchi, M. & Dzau, V. J. *Biochem. biophys. Res. Commun.* **197**, 393–399 (1993).
- Soriano, P., Montgomery, C., Geske, R. & Bradley, A. *Cell* **64**, 693–702 (1991).
- Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. *Proc. natn. Acad. Sci. U.S.A.* **90**, 8424–8428 (1993).
- Wood, S. A., Allen, N. D., Rossant, J., Auerbach, A. & Nagy, A. *Nature* **365**, 87–89 (1993).
- Xu, M. *et al.* *Cell* **79**, 945–955 (1994).

ACKNOWLEDGEMENTS. We thank A. Nagy for the R1 embryonic stem cell line, D. Bernstein for assistance with blood pressure measurements, and C. Hurt for help with statistical analysis. This work was supported in part by the NIH grants (R.E.P. and V.J.D.), and by the Howard Hughes Medical Institute (G.S.B. and B.K.K.). L.H. was supported by a fellowship from the Deutsche Forschungsgemeinschaft. V.J.D. is the recipient of an NIH MERIT award.

Effects on blood pressure and exploratory behaviour of mice lacking angiotensin II type-2 receptor

Toshihiro Ichiki, Patricia A. Labosky*, Chiyo Shiota, Shigeru Okuyama†, Yasuko Imagawa†, Agnes Fogo‡, Fumio Niimura‡, Iekuni Ichikawa‡, Brigid L. M. Hogan* & Tadashi Inagami§

Departments of Biochemistry, * Cell Biology and ‡ Pediatrics, and * Howard Hughes Medical Institute, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA

† First Laboratory, Medicinal Research Laboratories, Taisho Pharmaceutical Co. Ltd, No 403, Yoshino-cho 1-Chome, Ohmiya-shi, 330 Saitama, Japan

§ To whom correspondence should be addressed

THERE are two major angiotensin II receptor isoforms, AT₁ and AT₂. AT₁ mediates the well-known pressor and mitogenic effects of angiotensin II (refs 1–5), but the signalling mechanism and physiological role of AT₂ (refs 6–11) has not been established. Its abundant expression in fetal tissues¹² and certain brain nuclei¹³ suggest possible roles in growth, development and neuronal functions. Here we report the unexpected finding that the targeted disruption of the mouse AT₂ gene resulted in a significant increase in blood pressure and increased sensitivity to the pressor action of angiotensin II. Thus AT₂ mediates a depressor effect and antagonizes the AT₁-mediated pressor action of angiotensin II. In addition,

TABLE 1 Exploratory behaviour in F₂ wild-type male (control) and F₂ hemizygous (AT₂^{tm1ti}) mice

	Starting latency (s)	Ambulation (count)	Rearing (time)
Control (n=10)	9.3±1.9	60.9±5.9	12.9±1.8
AT ₂ ^{tm1ti} (n=11)	9.7±2.4	40.5±4.8*	9.1±2.5

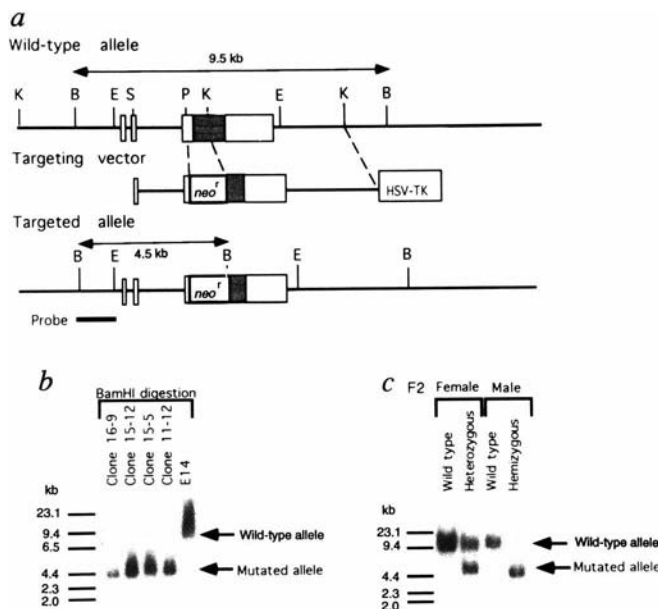
Mice were housed individually in acrylic cages. The apparatus used in the test has been described previously¹⁸. The floor of the open field (30×50 and 15 cm high) was divided into 15 identical squares (10×10 cm). Mice 13–15 weeks old were placed individually in one corner of the open field, and the time elapsed before they started exploring the environment was recorded as the starting latency. They were allowed to explore the environment freely for 3 min. During this period, ambulations were measured by counting the number of times that mice crossed from one square to another in the open field. The frequency of rearing was also counted. AT₂^{tm1ti} mice showed decreased ambulation and rearing activity. Dopamine-receptor antagonists have been reported to reduce ambulation by 10–30% by the same analysis¹⁸. Experimenters were not informed of the AT₂ genotype. Statistical analyses were performed by analysis of variance (ANOVA) followed by *post hoc* tests according to Tukey when appropriate. Data are expressed as means ± s.e.m. **P*<0.05 versus control.

disruption of the AT₂ gene attenuated exploratory behaviour and lowered body temperature. Our results show that angiotensin II activates AT₁ and AT₂, which have mutually counteracting haemodynamic effects, and that AT₂ regulates central nervous system functions, including behaviour.

We disrupted the mouse AT₂ gene, which contains the entire coding region in the third exon¹⁴ (Fig. 1a), in embryonic stem (ES) E14-1 (ref. 15) cells (Fig. 1b) and generated chimaeric mice (Fig. 1 legend). Because the AT₂ gene is on the X chromosome¹⁶ and ES cells are XY, all F₁ male mice were wild-type and all F₁

FIG. 1 Targeted disruption of the gene encoding the angiotensin II type-2 receptor (AT₂). a, Top, restriction map and exon-intron organization of the mouse AT₂ gene. Open boxes indicate exons and the shaded region shows the coding region of the mouse AT₂ gene. Middle, targeting vector used to disrupt the AT₂ gene. The PstI-KpnI segment was replaced with a neomycin-resistance expression cassette (*neo*^r). A herpes simplex virus thymidine kinase (HSV-TK) expression cassette added to the 3' end was used to eliminate clones having random integration of the targeting vector by Ganciclovir (Syntex Inc.). Bottom, restriction map and genomic structure of the mutated AT₂ gene AT₂^{tm1ti} (AT₂ targeted mutation 1 Tadashi Inagami or Toshihiro Ichiki, used as an identification or registry number) after homologous recombination. Restriction sites: K, KpnI; B, BamHI; E, EcoRI; S, SpeI; P, PstI. The BamHI-EcoRI segment corresponding to the region indicated as probe was used to screen for homologous recombinants. b, Southern blot analysis of genomic DNA of targeted ES clones and parental ES (E14-1) cells. Digestion with BamHI gave a 9.5-kb band for the wild-type allele and a 4.5-kb band for the mutated allele. Note that the wild-type allele is not present in AT₂-targeted clones. This is because the AT₂ gene is on the X chromosome and ES cells are derived from male blastocysts. c, Southern blot analysis of tail DNA of F₂ mice. Chimaeric males were crossed with C57BL/6 females. Of 112 F₂ mice produced from the cross of wild-type F₁ males and heterozygous F₁ females, 26 (23.2%) were wild-type females, 30 (26.8%) were heterozygous females, and 29 (25.9%) males were wild-type and 27 (24.1%) were hemizygous mutant males. The even distribution indicates that there was little or no loss of AT₂^{tm1ti} mice *in utero*.

METHODS. A genomic clone containing the mouse AT₂ gene was isolated from a genomic DNA library of 129/SV mouse (Stratagene). The SpeI-PstI fragment of the AT₂ gene, *neo*^r from pMC1neo^rpoly(A), the KpnI fragment of the AT₂ gene, and HSV-TK from PGK-tk were cloned into the multiple cloning sites of pBluescript SK(+) vector. 8×10⁷ of embryonic stem (ES) cells of E14-1 (ref. 15) line were electroporated with 200 µg of linearized targeting vector (single pulse of 2.1 kV and 500 µF in a 0.4-cm cuvette). Transfected cells were plated onto a feeder layer of irradiated embryonic *neo*^r fibroblasts and selected by G418 (0.3 mg ml⁻¹) and Ganciclovir (2 µM). About 800 double-resistant colonies were screened for homologous recombination by Southern blot



analysis²⁹. Four targeted clones were obtained and two clones (11-12 and 15-12) were injected into blastocysts derived from C57BL/6 mice and implanted into the uterus of pseudopregnant ICR mice. Out of 34 chimaeras, germline transmission occurred in 11 mice derived from both clones combined. Heterozygous female mice and wild-type male mice were mated to generate hemizygous mutant male mice. Hemizygous mutant male mice derived from both clones showed the same phenotype.

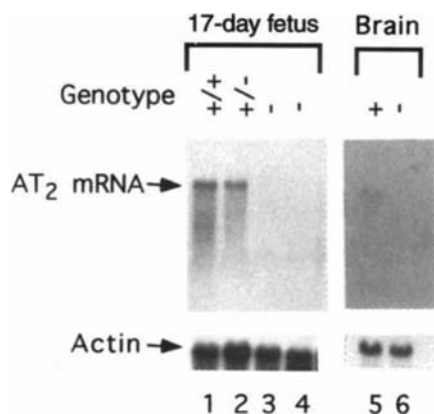


FIG. 2 Northern blot analysis of AT_2 mRNA. Poly(A)⁺ RNA was prepared from wild-type female (+/+), heterozygous female (+/-), hemizygous male (-/-) fetuses at 17 days p.c. and from the brain of wild-type (+) and hemizygous (-) male mice. Sex of fetuses was determined by the density and AT_2 gene genotype of Southern blot analysis. Poly(A)⁺ RNA (2 μ g) was electrophoresed in a formaldehyde gel, transferred to a nylon membrane. The blots were hybridized with mouse AT_2 cDNA, then rehybridized with β -actin cDNA to verify equal loading.

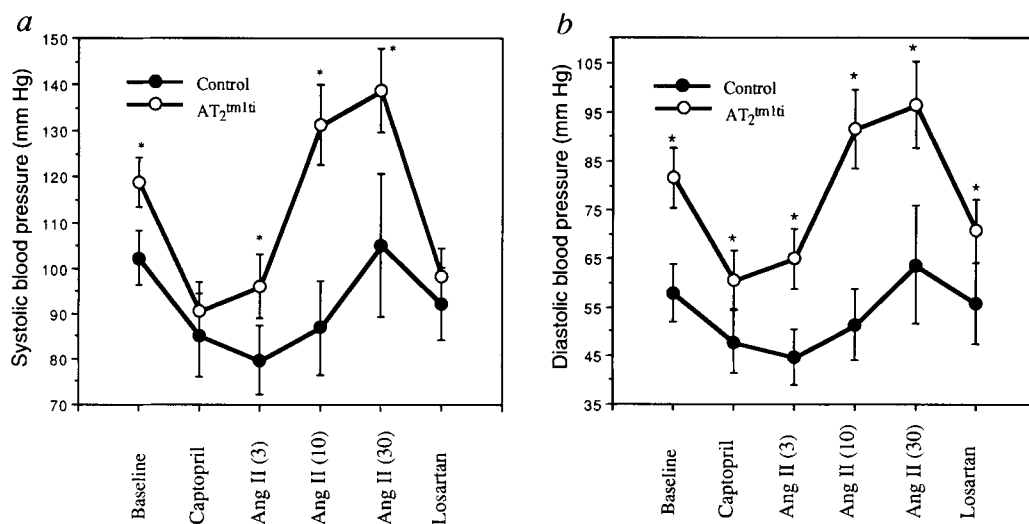
METHODS. Poly(A)⁺ RNA was prepared using a Fast Track kit (Invitrogen). 2 μ g poly(A)⁺ RNA was electrophoresed in a 1% agarose/1% formaldehyde gel and transferred to a nylon membrane by capillary transfer in 10 $\times$ SSC buffer (1 $\times$ SSC is 300 mM NaCl, 30 mM sodium citrate) overnight, baked for 2 h at 80 $^{\circ}$ C. Mouse AT_2 and β -actin cDNAs were labelled with 32 P using a Prime-it kit (Stratagene). Prehybridization, hybridization and washing have been described previously¹⁴.

females were heterozygous for the mutation. These heterozygous females, which appeared to be healthy, were mated with F₁ males, giving rise to an F₂ generation that included hemizygous males and heterozygous females (Fig. 1c). Homozygous females were obtained in the F₃ generation by mating these F₂ mice. The hemizygous males and homozygous females appeared to be healthy and were fertile. RNA from hemizygous males was analysed by northern blotting to confirm the absence of AT_2 transcripts in fetuses and in adult brains, both of which normally express AT_2 at high levels^{12,14} (Fig. 2). Levels of AT_1 mRNA were not affected in either kidney or brain (not shown). Hemizygous males were examined for histological abnormalities in regions that normally express AT_2 , including brain (inferior olive and locus coeruleus), heart, lung, kidney and adrenal gland; no obvious morphological changes were seen.

To determine the role of AT_2 in blood pressure (BP) regulation, we measured the basal BP level and its response to angio-

tensin II infusion in conscious mice by a direct method (Fig. 3). The baseline BP of F₂ hemizygous male mice was significantly higher than that of F₂ wild-type male mice (control). The difference was statistically significant in both systolic (SBP) and diastolic BP (DBP). Administration of an intravenous bolus of captopril, an angiotensin I converting enzyme inhibitor, to eliminate endogenous angiotensin II decreased BP both in the control and hemizygous mutant mice. DBP was significantly different between the control and hemizygous mutant mice after captopril administration. Upon infusion of angiotensin II, the increase in BP was greater in mutant than in control mice and the difference was statistically significant at all doses. Losartan, an AT_1 receptor-specific antagonist, reduced BP to the same extent as captopril administration in both hemizygous mutant and control mice. DBP after losartan administration was significantly different between these mice. The heart rate was generally higher in the mutants than in control mice at

FIG. 3 Systolic and diastolic blood pressure and their response to captopril, angiotensin II and losartan in F₂ wild-type male (control) and F₂ hemizygous male (AT_2^{m1ti}) mice. a, Systolic blood pressure (SBP), and b, diastolic blood pressure (DBP) in control male mice and AT_2^{m1ti} mice. After measuring baseline blood pressure (BP), captopril, three doses of angiotensin (Ang) II and losartan were intravenously administered in succession. Numbers in parentheses indicate the infusion rate of Ang II in ng per min per kg body weight. BP values between the baseline measurement and captopril treatment, between the captopril and Ang II treatment, and between losartan and captopril



treatment or Ang II (30 ng per kg per min) treatment were compared separately. Statistical analyses were performed by analysis of variance (ANOVA), followed by *post hoc* tests according to Bonferroni when appropriate. Data are expressed as means $\pm$ s.e.m. * P < 0.05 versus control. Six control and eight AT_2^{m1ti} mice derived from two ES clones were examined. Experimenters were not informed of the AT_2 genotype. METHODS. Mice weighing 35–40 g (12–16 weeks old) were anaesthetized with methoxyflurane (Pitman-Moore) and placed on a controlled-warming table. PE10 tubing, heated and tapered at one end, was filled with heparin (100 U ml⁻¹)-saline solution and inserted into the right femoral artery. The remaining portion was threaded under the skin and exited at the back of the nape, where the end of the cannula was sealed by heating. The second cannula using PE10 tubing was inserted into

the left jugular vein, threaded under the skin and also exited at the back of the nape. 24 h after surgery, a piece of PE50 tubing was connected to the femoral artery cannula. The other end of the tubing was connected to a swivel to allow free mobility of the mice. BP was measured in conscious mice with a Cobe CDX III transducer which was connected to a MicroMed BP Analyzer. BP and heart rate were continuously monitored during the following protocol. Baseline values were recorded for 10–30 min until stable. Mice were then given a bolus administration of captopril (30 mg per kg body weight) over a 30-s period in the jugular cannula. After 10 min of captopril administration, Ang II diluted in saline was infused at doses of 3, 10 and 30 ng per min per kg body weight for 5 min each. At the end of the Ang II infusion, losartan (10 mg per kg; DuPont Merck) was administered.

baseline and after each treatment, but the difference was not statistically significant.

Because AT_2 is expressed in the locus coeruleus¹³ and other brain nuclei, we examined the behaviour of hemizygous male mice by using a multiparametric check list¹⁷. Mutant males showed an increase in fear and hypoalgesia and a small decrease in startle response. Awareness, posture, motor coordination, muscle tone and reflexes were not affected. To test their exploratory behaviour, mice were placed individually in one corner of an open field which was divided into 15 identical squares¹⁸. The time elapsed before they started exploring the environment (starting latency), the number of times that mice crossed from one square to another (ambulation) and the number of rearing events, were measured (Table 1). Ambulation was markedly reduced in the mutant mice, suggesting that their exploratory behaviour was affected. The number of rearings was decreased, although the difference was not statistically significant.

Rectal temperature was measured using an electrothermometer after four hours of acclimatization to the environment. Mutant males showed a significant decrease in body temperature ($37.4 \pm 0.1^\circ\text{C}$ compared with $37.9 \pm 0.1^\circ\text{C}$ for wild-type controls; n was 10 and 11, respectively; $P < 0.05$).

The most striking effect of disrupting the AT_2 gene is the elevation of basal blood pressure and its increased sensitivity to angiotensin II. The difference in basal DBP persisted even when AT_1 receptors were blocked by losartan, indicating that this effect was independent of AT_1 . In contrast, the pressor response to infused angiotensin II requires AT_1 , and the fact that this response is greater in mutants than controls indicates that AT_2 normally serves to limit the response of the AT_1 receptor to angiotensin II. This anti-hypertensive action was unexpected, because it is generally thought that angiotensin II is a pressor factor, and mice lacking the angiotensinogen gene show reduced BP^{19,20}. Our results, however, indicate that angiotensin II acts through two receptors with opposing effects, and suggest that BP *in vivo* may be determined by the balance between AT_1 and AT_2 activation. The exact mechanism is likely to be complex, and the effect of the AT_2 -specific antagonist PD12319 on BP is controversial^{21–25}. More investigation is therefore required to understand the anti-hypertensive action of AT_2 . One limitation of our study is that we cannot exclude the possibility that strain-129-derived alleles of other gene(s) on the X chromosome affect the phenotypic difference we found between mutant male and control mice.

Hein *et al.* also deleted the AT_2 gene²⁶. Like us, they observed increased sensitivity to angiotensin II in mutant mice, but in contrast to our results, they did not see any increase in basal BP. This could be due to differences in genetic background: although both groups used ES cells from strain 129 mice, we crossed the chimaeric males with C57BL/6 females, whereas they used FVB/N females. Our angiotensin II administration protocol also differed from theirs, making direct comparison difficult. We considered the possibility that the increase in BP might be a consequence of surgical stress, but this is unlikely because the difference was still observed when we measured BP under inactin anaesthesia, which has a minimal cardiovascular-suppressive effect on mice. We found a significant increase in both SBP (118.2 ± 5.0 mm Hg in male mutant mice ($n = 6$) compared with 94.2 ± 1.7 mm Hg in control male mice ($n = 6$), $P < 0.01$) and DBP (86.2 ± 3.1 mm Hg compared with 66.0 ± 2.1 mm Hg in controls, $P < 0.01$) in male mutant mice.

Although it is known that the renin-angiotensin system is present in the brain²⁷, its effect on behaviour has not been clearly defined. AT_2 is expressed in the locus coeruleus, which supplies a noradrenergic projection to many brain areas and is thought to regulate arousal, attention, anxiety, and affective and pain responses²⁸. Our finding that mice lacking AT_2 show reduced exploration, increased fear, hypoalgesia and lower body temperature warrants further neurophysiological analysis. □

Received 24 April; accepted 15 September 1995.

1. Peach, M. J. *Biochem. Pharmacol.* **30**, 2745–2751 (1981).
2. Bottari, S. P., de Gasparo, M., Steckelings, U. M. & Levens, N. R. *Front. Neuroendocr.* **14**, 123–171 (1993).
3. Timmermans, P. B. M. W. M. *et al. Pharmac. Rev.* **45**, 205–251 (1993).
4. Sasaki, K. *et al. Nature* **351**, 230–233 (1991).
5. Murphy, T. J. *et al. Nature* **351**, 233–236 (1991).
6. Kambayashi, Y. *et al. J. Biol. Chem.* **268**, 24543–24546 (1993).
7. Mukoyama, M. *et al. J. Biol. Chem.* **268**, 24539–24542 (1993).
8. Summers, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 7567–7571 (1991).
9. Buisson, B. *et al. J. Biol. Chem.* **270**, 1670–1674 (1995).
10. Stoll, M. *et al. J. clin. Invest.* **95**, 651–657 (1995).
11. Lo, M. *et al. J. clin. Invest.* **95**, 1394–1397 (1995).
12. Grady, E. F. *et al. J. clin. Invest.* **88**, 921–933 (1991).
13. Tsutsumi, K. & Saavedra, J. M. *Am. J. Physiol.* **261**, R209–R216 (1991).
14. Ichiki, T. & Inagami, T. *Circ. Res.* **76**, 693–700 (1995).
15. Kuhn, R., Rajewsky, K. & Muller, W. *Science* **254**, 707–710 (1991).
16. Koike, G. *et al. Biochem. biophys. Res. Commun.* **203**, 1842–1850 (1994).
17. Irwin, S. in *Drug Screening and Evaluation of New Compounds in Animals* (ed. Nodine, J. H. & Siegler, D. E.) 36–54 (Year Book Medical, Chicago, 1964).
18. Ichihara, K., Nabeshima, T. & Kameyama, T. *J. Pharmac. Exp. Ther.* **264**, 122–128 (1993).
19. Tanimoto, K. *et al. J. Biol. Chem.* **269**, 31334–31337 (1994).
20. Smithies, O. & Kim, H. S. *Proc. natn. Acad. Sci. U.S.A.* **91**, 3612–3615 (1994).
21. Tofovic, S. P., Pong, A. S. & Jackson, E. K. *Hypertension* **18**, 774–782 (1991).
22. Keiser, J. A. *et al. J. Pharmac. Exp. Ther.* **262**, 1154–1160 (1992).
23. Naver, L., Stromberg, C. & Saavedra, J. M. *Reg. Peptide* **52**, 21–29 (1994).
24. Scheuer, D. A. & Perrone, M. H. *Am. J. Physiol.* **264**, R917–R923 (1993).
25. Yoshida, M., Hamakubo, T. & Inagami, T. *Am. J. Physiol.* **266**, R802–R808 (1994).
26. Hein, L., Barsh, G. S., Richard, R. E., Dzau, V. J. & Kobilka, B. K. *et al. Nature* **377**, 744–748 (1995).
27. Phillips, M. I., Speakman, E. A. & Kimura, B. *Reg. Peptides* **43**, 1–20 (1993).
28. Ashton, H. in *Brain Systems, Disorders, and Psychotropic Drugs* (ed. Ashton, H.) 1–49 (Oxford University Press, 1987).
29. Solis-Ramirez, R. *et al. in Guide to Techniques in Mouse Development* (eds Wassarman, P. M. & DePamphilis, M. L.) 855–878 (Academic, San Diego, 1993).

ACKNOWLEDGEMENTS. We thank T. Watanabe and D. Kitamura for advice on the gene targeting technique and for E14-1 ES cells, and E. Price Jr for technical assistance.

Activation of the SAPK pathway by the human *STE20* homologue germinal centre kinase

Celia M. Pombo*, John H. Kehrl†, Irma Sánchez, Paul Katz‡, Joseph Avruch, Leonard I. Zon§, James R. Woodgett||, Thomas Force* & John M. Kyriakis*

Diabetes Research Laboratory and * Cardiology Unit, Massachusetts General Hospital East and Department of Medicine, Harvard Medical School, 149 13th Street, Charlestown, Massachusetts 02129, USA

† Laboratory of Immunoregulation, NIAID, National Institutes of Health, Building 10, 10 Center Drive MSC 1876, Bethesda, Maryland 20892, USA

‡ Department of Medicine, Georgetown University Medical Center, Washington, DC 20007, USA

§ Howard Hughes Medical Institute and Division of Hematology/Oncology, Children's Hospital and Dana Farber Cancer Institute, Harvard Medical School, Enders 650, 300 Longwood Avenue, Boston, Massachusetts 02115, USA

|| Ontario Cancer Institute, Princess Margaret Hospital, 500 Sherbourne Street, Toronto, Ontario M4X 1K9, Canada

EUKARYOTIC cells respond to different extracellular stimuli by recruiting homologous signalling pathways that use members of the MEKK, MEK and ERK families of protein kinases. The MEKK→MEK→ERK core pathways of *Saccharomyces cerevisiae* may themselves be regulated by members of the *STE20* family of protein kinases^{1,2}. Here we report specific activation of the mammalian stress-activated protein kinase (SAPK) pathway by germinal centre kinase (GCK; ref. 3), a human *STE20* homologue^{3,4}. SAPKs, members of the ERK family, are activated

* To whom correspondence should be addressed.

in situ by inflammatory stimuli, including tumour-necrosis factor (TNF) and interleukin-1, and phosphorylate and probably stimulate the transactivation function of c-Jun⁵⁻⁷. Although GCK is found in many tissues, its expression in lymphoid follicles is restricted to the cells of the germinal centre, where it may participate in B-cell differentiation³. Activation of the SAPK pathway by GCK illustrates further the striking conservation of eukaryotic signalling mechanisms and defines the first physiological function of a mammalian Ste20.

Resting 293 (Fig. 1a) or COS (Fig. 1b) cells coexpressing haemagglutinin (HA)-tagged SAPK and GCK show almost 15-fold activation of SAPK. Anisomycin, a good SAPK agonist^{5,7}, activated the 293 cell SAPKs only fourfold in the absence of

coexpressed GCK. Anisomycin treatment of cells coexpressing SAPK and GCK resulted in no further stimulation of 293 cell SAPK (Fig. 1a) but gave a further twofold activation of COS cell SAPK activity (Fig. 1b). Thus GCK-independent mechanisms of SAPK activation probably exist. Indeed, anisomycin does not activate endogenous B-cell GCK activity (see below). Coexpression of SAPK and GCK resulted in no detectable changes in SAPK expression levels (Fig. 1c).

GCK displayed a distinct selectivity towards downstream targets, activating neither p44-MAP kinase (Fig. 2b) nor p38 (Fig. 2c), a mammalian homologue of the *S. cerevisiae* osmosensing ERK, *HOG1* (ref. 8), under conditions wherein SAPK is activated nearly fivefold (Fig. 2a). GCK had no effect on p44-MAPK or p38, in spite of the fact that both the MAPK and p38 pathways remained intact and could be activated, respectively, by coexpression with oncogenic Raf-1 or by hyperosmolarity (Fig. 2b, c). These other mammalian ERKs participate in distinct signalling mechanisms. p44 MAPK is critical to the cellular response to mitogen^{9,11}. p38, like the SAPKs, is activated by cellular stress and inflammatory cytokines^{8,12,13}. However, p38

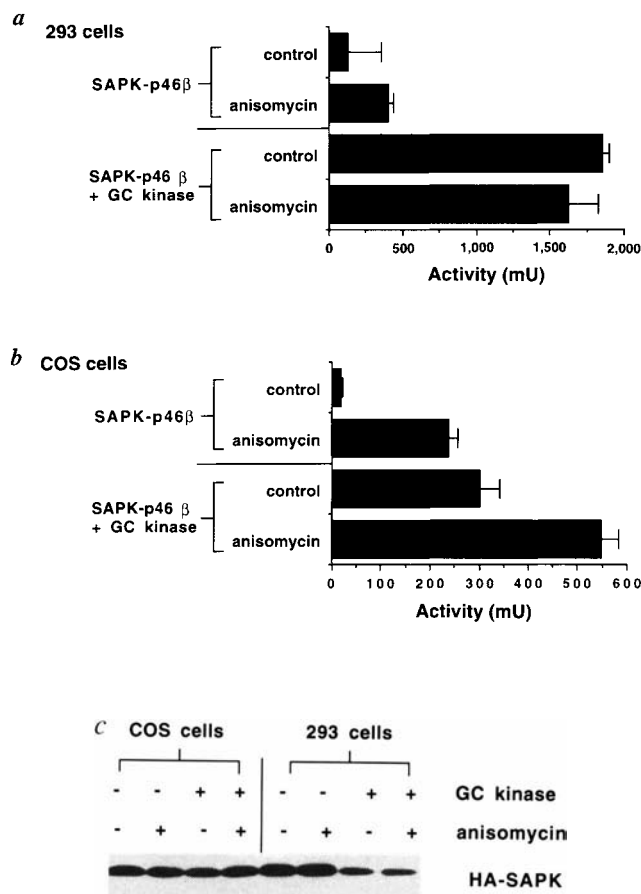


FIG. 1 Coexpression with GCK activates SAPK. **a**, 293 cells were transfected with either HA-SAPK or HA-SAPK and GCK as indicated. After 48 h, cells were treated with anisomycin and HA-SAPK was immunoprecipitated from cell lysates and assayed for kinase activity in immune complexes. Glutathione-S-transferase-c-Jun was the substrate. **b**, As **a**, except that COS cells were used. **c**, Coexpression with GCK has no effect on HA-SAPK levels. Portions of extracts from cells transfected with HA-SAPK and GCK were immunoblotted with anti-HA antibody. **METHODS**. 293 cells were transfected using the calcium phosphate method⁷. COS cells were transfected using DEAE-dextran as described³. In both cases, each 10-cm plate was transfected with 10 μ g pMT3-SAPK-p46 β , encoding an HA-tagged construct of the 46K variant of SAPK- β , and 10 μ g pRc/CMV-GCK encoding full-length, unmodified GCK. Cells not transfected with GCK were transfected with an equivalent amount of empty plasmid. After 48 h, cells were treated with anisomycin (10 μ g ml⁻¹ for 15 min). Cell lysis, anti-HA immunoprecipitation and SAPK immune-complex kinase assay were performed as described⁵. Mean $\pm$ s.d. for triplicate assays are shown. 1 U SAPK will transfer 1 pmol min⁻¹ phosphate from ATP to GST-c-Jun (ref. 5). HA immunoblotting was performed using the enhanced chemiluminescence system (Amersham) according to the manufacturer's instructions.

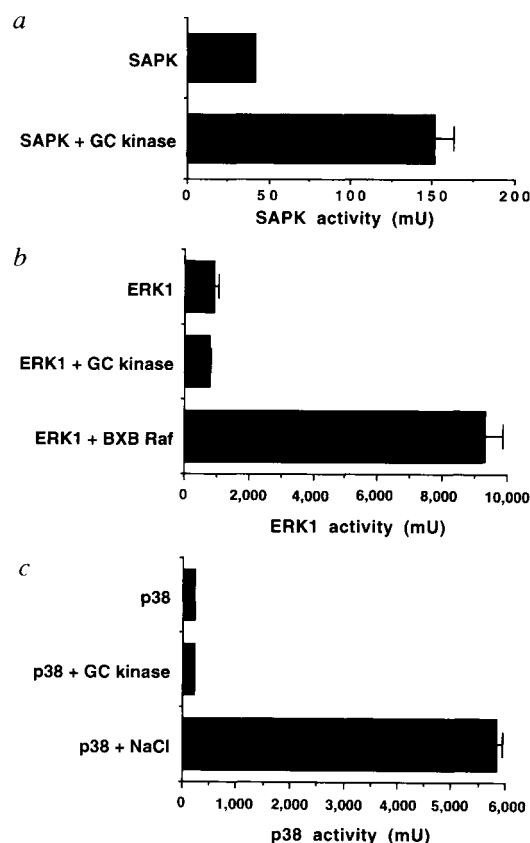


FIG. 2 GCK specifically activates SAPK, but not MAPK (ERK1) or the mammalian *HOG1* homologue, p38. **a**, Activation of resting-cell SAPK activity upon coexpression of low levels of GCK. **b**, p44 MAPK (ERK1) is not activated upon coexpression with low levels of GCK but is activated by low levels of oncogenic BXB Raf-1 (ref. 22). **c**, p38 is not activated by GCK but can be activated by hypertonic stress. **METHODS**. COS cells were transfected, as indicated, with 10 μ g pMT3 constructs encoding HA-SAPK-p46 β , HA-p44 MAPK (ERK1) or HA-p38, and 2.5 μ g pRc/CMV-GCK, pRSV-BXB-Raf-1 or the cognate empty plasmids as described in Fig. 1 legend. A portion of the cells transfected with only p38 was treated with hypertonic stress (0.7 M NaCl for 10 min) as indicated. After 48 h, cells were lysed and immunoprecipitated with anti-HA⁵. SAPK was assayed as for Fig. 1. MAPK and p38 were assayed for MBP kinase activity as described^{5,23}. Means $\pm$ s.d. for triplicate determinations are shown. 1 U of MAPK or of p38 will transfer 1 pmol min⁻¹ phosphate from ATP to MBP.

can be regulated both by a MEK upstream of the SAPKs (SEK1) or by MKK3, a MEK specific for p38 (refs 12, 14). Inasmuch as the SAPK and p38 pathways are activated by essentially identical extracellular stimuli, the divergence between the p38 and SAPK pathways may occur upstream of the level of Ste20s. Alternatively, as has been suggested for yeast, GCK may specifically target SAPK- but not p38-containing signalling modules held together by specific scaffolding proteins^{1,2,4,14}. We have also observed that GCK does not activate coexpressed p70 S6 kinase (K. Andrabi and J.A., unpublished results). Thus GCK is not a nonspecific stress/mitogenic activator and seems to activate the SAPK pathway preferentially.

SEK1 is a novel MEK which can activate the SAPKs *in vitro* and *in situ*^{7,15}. GCK activated both SEK1-catalysed SAPK phosphorylation (Fig. 3b) and activation (Fig. 3a). GST-SEK1 fusion protein from untreated 293 cells that did not express GCK activated SAPK threefold above background. Anisomycin treatment enhanced this activation to tenfold. By contrast, GST-SEK1 isolated from untreated cells coexpressing GCK activated SAPK to the extent (tenfold) observed for anisomycin-stimulated cells expressing GST-SEK1 alone; and anisomycin gave no further GST-SEK1 activation (Fig. 3a). SEK1-AL, a mutant SEK1 polypeptide in which the sites of regulatory phosphorylation (Ser 220, Thr 224) have been mutated to Ala and Leu, respectively¹⁶, can act as a dominant inhibitor of SAPK activation¹⁶. When expressed in COS cells, SEK1-AL can potentially inhibit GCK activation of the SAPKs (Fig. 3c). Taken together, these results demonstrate that GCK can activate the SAPKs through activation of SEK1.

Two classes of STE20 kinase have been described. Ste20p, a component of the *S. cerevisiae* pheromone-response pathway, consists of an amino-terminal regulatory domain and a carboxy-terminal catalytic domain⁴. p65^{PAK1}, a mammalian STE20 homologue, has a domain structure similar to that of Ste20 and is activated upon the interaction of its amino-terminal regulatory domain with Rac and Cdc42 (refs 17, 18). *SPS1* encodes a *S.*

cerevisiae STE20 homologue required for the pathway regulating sporulation¹⁹. Both Sps1 and GCK consist of an amino-terminal catalytic domain and a carboxy-terminal putative regulatory domain. Because overexpression of any mammalian STE20 with a domain structure similar to that of GCK might activate the SAPKs nonspecifically, and because the ligands that activate GCK are unknown, we investigated the specificity and relevance of SAPK activation by GCK. Accordingly, we examined the ability of another mammalian STE20 homologue, similar in overall structure to GCK, to activate the SAPKs. In addition, we assayed endogenous GCK activated *in situ* by a variety of ligands known to activate the SAPKs (ref. 5). Upstream kinase-1 (UK1) is a novel mammalian STE20 which, like GCK (ref. 3), is highly expressed in B cells (C.M.P., J.M.K. and T.F., manuscript in preparation); the overall domain structure of UK1 is strikingly similar to that of GCK: the UK1 and GCK catalytic domains share >45% identity and both kinases have an extended carboxy-terminal non-catalytic domain. Like GCK (ref. 3), overexpressed HA-tagged UK1 is constitutively active towards myelin basic protein (MBP; Fig. 4a, left, lanes 9 and 10). However, UK1 cannot activate coexpressed GST-SAPK under conditions in which coexpressed GCK can (Fig. 4a, left, lanes 1–6, and right). Thus overexpression of a mammalian STE20 does not result in promiscuous SAPK activation, and, in fact, the mammalian STE20s display remarkable specificity (Figs 2 and 4a). Still, it may be possible for the SAPKs to receive activating input from multiple upstream components, including other members of the STE20 family. Indeed, recent data have implicated Rac, Cdc42 and p65^{PAK} in the activation of the SAPKs (refs 20, 21).

Anti-GCK antibodies were used to immunoprecipitate endogenous GCK from RAMOS cells, a human Burkitt lymphoma line. Resting GCK activity, measured using MBP as a substrate, is high in these cells, and is further stimulated *in situ* by TNF- α , but not by Fas or anisomycin (Fig. 4b). Insofar as TNF- α is a potent SAPK agonist in many cell types, including B cells⁵

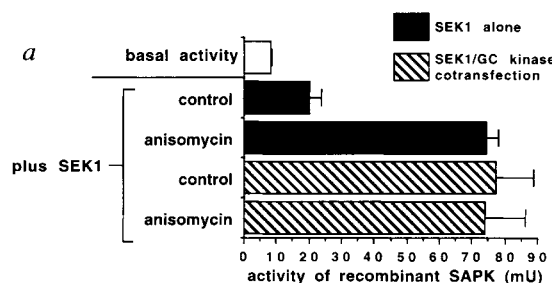
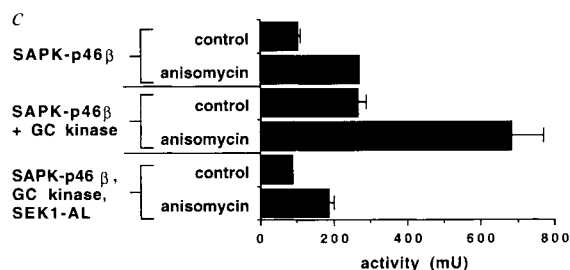
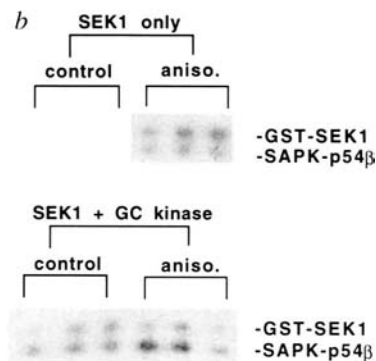


FIG. 3 Activation of SEK1 by GCK. **a**, Activation of SEK1 upon coexpression with GCK. GST-SEK1 was assayed for activation of recombinant SAPK *in vitro*. The basal activity of the SAPK preparation is indicated by the white bar. Activity of the SAPK preparation after incubation with GST-SEK1 from vehicle or anisomycin-treated 293 cells expressing GST-SEK1 only (black bars) or from cells expressing GCK and GST-SEK1 (striped bars) is indicated. **b**, GCK coexpression enhances SEK1-catalysed phosphorylation of the SAPK polypeptide. The positions of the SAPK and GST-SEK1 polypeptides are indicated, as are transfections (SEK1 or SEK1 plus GCK) and treatments (vehicle or anisomycin). **c**, Dominant inhibitor SEK1 (SEK-AL) inhibits GCK activation of SAPKs. METHODS. 293 cells were transfected with 10 μ g pEBG-SEK1 encoding GST-SEK1 (ref. 7) and either 10 μ g pRc/CMV-GCK or empty plasmid. For **c**, COS cells were transfected with 10 μ g pMT3-SAPK-p46 β , and either 6 μ g pRc/CMV-GCK, 6 μ g pRc/CMV-GCK plus 10 μ g pcDNA3-SEK1-AL (encoding a non-phosphorylatable, dominant inhibitory SEK1 construct¹⁶), or the cognate empty plasmids. Cells were then treated with anisomycin as described (Fig. 1 legend). For **a**, SEK1 was purified from cell lysates by glutathione-agarose chromatography and assayed for its ability to activate SAPK-p54 α as described⁷. Means $\pm$ s.d. for triplicate determinations are shown. For **b**, SEK1-catalysed phosphorylation of the SAPK polypeptide was analysed after SDS-PAGE as in **a**. For **c**, SAPK was immunoprecipitated from the COS cells and assayed as described (Fig. 1 legend). Means $\pm$ s.d. for triplicate determinations are shown.



ation of the SAPK polypeptide was analysed after SDS-PAGE as in **a**. For **c**, SAPK was immunoprecipitated from the COS cells and assayed as described (Fig. 1 legend). Means $\pm$ s.d. for triplicate determinations are shown.

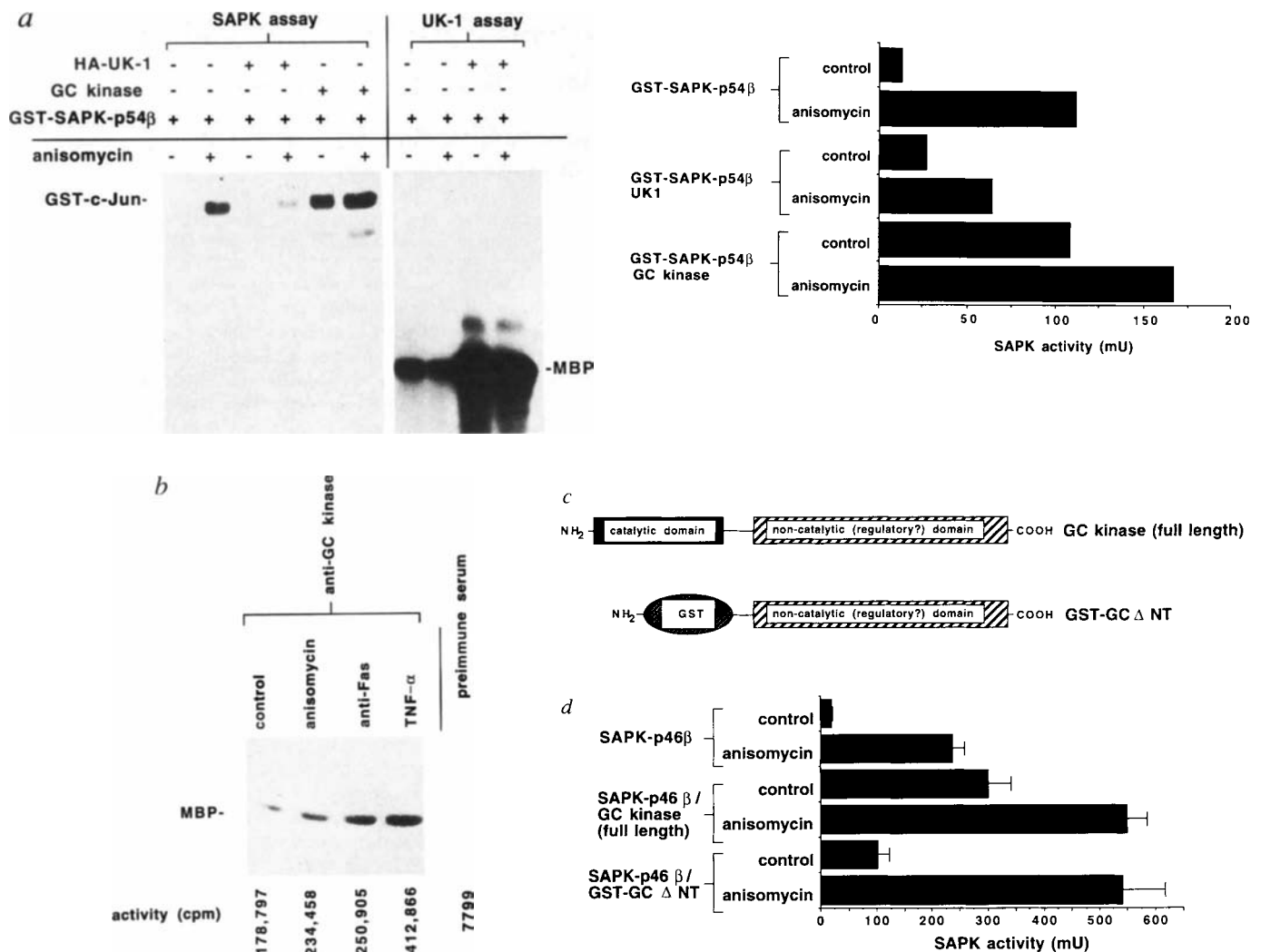


FIG. 4 Specificity and regulatory features of GCK activation of SAPK. **a**, Specificity of mammalian STE20 signalling. UK1, a novel mammalian STE20 does not activate the SAPKs under conditions wherein GCK does. Left, SAPKs are activated by anisomycin (lanes 1 and 2, indicated) or upon coexpression with GCK (lanes 5 and 6) but not by UK1 (lanes 3 and 4). Lanes 9 and 10 are assays of HA-UK1 taken from cells coexpressing GST-SAPK and UK1. Lanes 7 and 8 are parallel assays of cells expressing SAPK and empty pCMV. Right panel, bars represent assays of SAPK; means of duplicate assays are shown. **b**, Endogenous GCK is activated by tumour-necrosis factor TNF- α in RAMOS cells. RAMOS cells were treated with the ligands indicated and GCK was immunoprecipitated and assayed for MBP kinase activity. **c**, The GCK GC Δ NT construct. **d**, Coexpression of SAPK and the C-terminal non-catalytic domain of GCK (GC Δ NT) results in SAPK activation. **e**, Neither GC kinase nor GC Δ NT has any effect on the levels of coexpressed HA-SAPK. HA SAPK was expressed alone, with full-length GCK, or with GC Δ NT and cells were treated with anisomycin, as indicated. A portion of each extract was assayed for SAPK activity and another portion was immunoblotted for SAPK expression with anti-HA antibody.

METHODS. To generate pEBG-GC Δ NT, the following strategy was used. A 1,604-bp cDNA encoding the non-catalytic C terminus of GCK was isolated as a PCR fragment with *Bgl*II and *Not*I sites at the 5' and 3' ends, respectively. The sense primer used was GGAGATCTCTGCTGGACAAAGCCAGTGACCCTCATCTG, the *Bgl*II site is underlined. The antisense primer used was AAGGAAAAAGCGGCCGCTAGGTGCTCTGGTGGCCCG, the *Not*I site is underlined. The fragment was cut with *Bgl*II and *Not*I and cloned, in frame, into the *Bam*HI and *Not*I sites of pEBG (ref. 7).

COS cells were transfected with pMT3-SAPK-p46 β and either pEBG-GC Δ NT or empty pEBG. After 48 h, cells were treated with anisomycin, lysed and assayed for activation of SAPKs as described (Fig. 1 legend). Means $\pm$ s.d. for triplicate determinations are shown. Immunoblotting was performed as described (Fig. 1 legend). For **a**, a fragment of ~400 bp of UK1 was amplified by PCR using degenerate primers based on known protein kinase sequences, and used to isolate the full-length clone from a human B-cell cDNA library. The methods for cloning, sequence and function of UK1 will be presented elsewhere (C.M.P., J.M.K. and T.F., in preparation). UK1 was subcloned into pCMV with an HA tag. COS cells were transfected with either pCMV-HA-UK1, pRc/CMV-GCK, or pCMV and pEBG-SAPK-p54 β , which encodes GST-SAPK-p54 β . Extracts were prepared as described^{5,7}, split and GST-SAPK or HA-UK1 purified as described^{3,5,7}. For **b**, RAMOS cells were treated with anisomycin (10 μ g ml⁻¹, 20 min) anti-Fas (200 ng ml⁻¹, 20 min) or TNF- α (50 ng ml⁻¹, 20 min). GCK was immunoprecipitated and assayed for MBP kinase as described³.

(where it may mediate cell selection), GCK may, at least in part, transduce the TNF- α signal to the SAPKs.

The data in Figs 1–3 as well as direct assays³ indicate that overexpressed GCK is constitutively active. Thus endogenous GCK may be regulated *in situ* by oligomerization, or by limiting concentrations of an inhibitor, the effects of which are relieved by upstream stimuli—processes possibly mediated by the non-catalytic GCK carboxy terminus. Thus overexpression of the GCK carboxy terminus might activate the SAPKs either by promoting GCK oligomerization or competing away an inhibitor. To begin to test this hypothesis, we generated pEBG-GCANT, a construct consisting of the GCK carboxy terminus fused to GST (Fig. 4c). Coexpression of pEBG-GCANT and SAPK resulted in a fivefold activation of SAPK. Anisomycin treatment of GCANT/SAPK coexpressing cells resulted in a further fivefold stimulation of SAPK activity (Fig. 4d). Expression of GCANT had no effect on the levels of coexpressed HA-SAPK (Fig. 4e). Activation of the SAPKs by the GCANT construct was nearly half that incurred by wild-type GCK. Thus the GCK carboxy terminus may regulate GCK activation by oligomerization or through interaction with an inhibitor.

While it is likely that multiple mechanisms of SAPK activation exist, our results show that, as in yeast^{1,2,4,17,18}, a mammalian member of the *STE20* family of protein kinases, GCK, can specifically regulate the SAPK pathway, an ERK-based signalling cascade. These results imply an astonishing degree of evolutionary conservation in signal transduction pathways. The activation of the SAPKs by GCK is specific inasmuch as GCK fails to activate other mammalian ERK pathways, and not all mammalian Ste20s activate the SAPKs. Our results also indicate that, like the SAPKs, GCK is activated *in situ* by TNF- α , a potent SAPK agonist⁵, and that GCK activation may involve oligomerization or dissociation of an inhibitor. Activated GCK may in turn recruit the SAPKs in response to TNF- α . Although yeast genetics place the *STE20*s upstream of the MEKKs, the actual substrates of GCK are unknown. Our preliminary data indicate that GCK does not phosphorylate SEK1 directly. Moreover, we do not detect activation by GCK and MEKK1, an immediate upstream activator of SEK1 (ref. 16). The prominent expression of GCK in the follicular germinal centre suggests that the GCK/SAPK pathway may be important in the B-cell differentiation and selection that occur in the germinal centre. □

Received 18 April; accepted 7 September 1995.

- Levin, D. E. & Errede, B. *Curr. Opin. Cell Biol.* **7**, 197–202 (1995).
- Herskowitz, I. *Cell* **80**, 187–197 (1995).
- Katz, P., Whalen, G. & Kehrl, J. H. *J. Biol. Chem.* **269**, 16802–16809 (1994).
- Leberer, E. et al. *EMBO J.* **11**, 4815–4824 (1992).
- Kyriakis, J. M. et al. *Nature* **369**, 156–160 (1994).
- Dérjard, B. et al. *Cell* **76**, 1025–1037 (1994).
- Sánchez, I. et al. *Nature* **372**, 794–798 (1994).
- Han, J., Lee, J.-D., Bibbs, L. & Ulevitch, R. J. *Science* **265**, 808–811 (1994).
- Avruch, J., Zhang, X.-f. & Kyriakis, J. M. *Trends biochem. Sci.* **19**, 279–283 (1994).
- Cowley, S., Paterson, H., Kemp, P. & Marshall, C. J. *Cell* **77**, 841–852 (1994).
- Mansour, S. J. et al. *Science* **265**, 966–970 (1994).
- Rouse, J. et al. *Cell* **78**, 1027–1037 (1994).
- Freshney, N. W. et al. *Cell* **78**, 1039–1049 (1994).
- Choi, K.-Y., Scatterberg, B., Lyons, D. M. & Ellison, E. A. *Cell* **78**, 499–512 (1994).
- Dérjard, B. et al. *Science* **267**, 682–685 (1995).
- Yan, M. et al. *Nature* **372**, 798–800 (1994).
- Manser, E. et al. *Nature* **367**, 40–46 (1994).
- Martin, G. A., Bollag, G., McCormick, F. & Abo, A. *EMBO J.* **14**, 1970–1978 (1995).
- Freisen, H., Lunz, R., Doyle, S. & Segall, J. *Genes Dev.* **8**, 2162–2175 (1994).
- Coso, O. A. et al. *Cell* **81**, 1137–1146 (1995).
- Minden, A. et al. *Cell* **81**, 1147–1157 (1995).
- Bruder, J. T., Heidecker, G. & Rapp, U. R. *Genes Dev.* **6**, 545–556 (1992).
- Lee, J. C. et al. *Nature* **372**, 739–746 (1994).

ACKNOWLEDGEMENTS. C.M.P. and J.H.K. contributed equally to this work. We thank A. S. Zervos for pMT3-p38, U. R. Rapp for pRSV-BXB-Raf-1, K. Andradi for preliminary experiments, G. Whalen for assistance and K. Dellovo for photography. J.M.K. is supported by grants from the USPHS and the US Army Breast Cancer Research Program. T.F. is supported by the USPHS and by a Grant-in-Aid from the American Heart Association. I.S. is supported by a USPHS Training Grant awarded to the M.G.H., J.H.K. and P.K. are supported by the Arthritis Foundation and an Interagency Personnel Agreement from the NIAID.

Initial hydrophobic collapse in the folding of barstar

Vishwas R. Agashe, M. C. R. Shastri
& Jayant B. Udgaonkar*

National Centre for Biological Sciences, TIFR Centre, PO Box 1234,
Indian Institute of Science Campus, Bangalore 560012, India

Two models are commonly used to describe the poorly understood earliest steps of protein folding. The framework model^{1–3} stresses very early formation of nascent secondary structures, which coalesce into a compact, molten, globule-like form^{4,5} from which tertiary structure slowly develops^{6,7}. The hydrophobic collapse model^{8–10} gives overriding precedence to a nonspecific collapse of the polypeptide chain which facilitates subsequent formation of specific secondary and tertiary structure^{11,12}. Here we report our analysis of the earliest observable events of the major folding pathway of barstar, a small protein. We compare the kinetics of folding using circular dichroism at 222 nm and 270 nm, intrinsic tryptophan fluorescence, fluorescence of the hydrophobic dye 8-anilino-1-naphthalene-sulphonic acid on binding, and restoration of tryptophan-dansyl fluorescence energy transfer as structure-monitoring probes. We show that the polypeptide chain rapidly collapses (within 4 ms) to a compact globule with a solvent-accessible hydrophobic core⁶, but with no optically active secondary or tertiary structure. Thus the earliest event of the major folding pathway of barstar is a nonspecific hydrophobic collapse that does not involve concomitant secondary structure formation.

Rapid-mixing experiments have demonstrated that a large fraction of the final secondary structure forms in a burst phase (<10 ms)^{7,13–18} of the folding pathways of all proteins studied so far. In a few cases, the secondary structure in the intermediate produced in the burst phase appears to be fluctuating and not stable^{15,19,20}, but it is nevertheless authentic²¹. Direct measurement of compactness of burst-phase kinetic intermediates has not yet been achieved²², but indirect measurements indicate that they are compact^{5,19,23,24} and possess sufficiently large hydrophobic clusters, formed by the collapse of non-polar residues onto each other, for binding of the hydrophobic dye 8-anilino-1-naphthalene-sulphonic acid (ANS)^{6,7,16,25,26}. It has not yet been possible to ascertain, for any protein studied, whether the initial collapse of the polypeptide chain precedes or follows secondary structure formation.

The folding pathway of barstar, the inhibitor of barnase in the bacterium *Bacillus amyloliquefaciens*, has been extensively characterized^{27–30}. Equilibrium-unfolded barstar comprises 69% slow-refolding (U_S) and 31% fast-refolding (U_F) molecules. The major folding pathway of barstar is $U_S \rightarrow I_{M1} \rightarrow I_{S1} \rightarrow I_N \rightarrow N$, where I_{M1} , I_{S1} and I_N are kinetic intermediates and N is the fully folded protein. In 1.0 M guanidine hydrochloride (GdnHCl), all U_S molecules fold by this pathway, and U_F molecules fold directly to N ³⁰. I_{M1} is a burst-phase intermediate; I_{S1} is an early intermediate with some tertiary structure because its fluorescence properties are different from those of U_S and I_{M1} ; and I_N is a native-like intermediate capable, like N , of inhibiting barnase, and its transition to N probably involves *trans-cis* isomerization of the Tyr47–Pro48 peptide bond²⁷. I_{M1} , I_{S1} and I_N all possess solvent-accessible hydrophobic cores on account of their ability to bind the hydrophobic dye ANS³⁰.

Here we have used five different spectroscopic probes to determine the chronological hierarchy of the earliest events of the major folding pathway of barstar. Folding has been performed in 1 M GdnHCl because the stabilities and therefore extents of accumulation of I_{M1} , I_{S1} and I_N are maximal at this denaturant concentration³⁰.

* To whom correspondence should be addressed.

TABLE 1 Kinetics of refolding of barstar in 1.0 M GdnHCl, pH 7

Spectroscopic probe	Burst-phase amplitude	λ_1 ($\times 10^{-3} \text{ s}^{-1}$)	λ_2 (s^{-1})	α_1	α_2
Ellipticity at 222 nm	0 $\pm$ 0.12	7 $\pm$ 3	11 $\pm$ 4	0.3 $\pm$ 0.1	0.7 $\pm$ 0.1
Ellipticity at 270 nm	0 $\pm$ 0.12	4 $\pm$ 2	4 $\pm$ 2	0.5 $\pm$ 0.1	0.5 $\pm$ 0.1
Intrinsic trp fluorescence	0 $\pm$ 0.12	9 $\pm$ 3	20 $\pm$ 4	0.3 $\pm$ 0.1	0.7 $\pm$ 0.1
Fluorescence change on ANS binding	1.0				
Restoration of trp-dansyl fluorescence energy transfer	0.7 $\pm$ 0.1	15 $\pm$ 3	7 $\pm$ 3	0.06 $\pm$ 0.01	0.24 $\pm$ 0.05

λ_1 and λ_2 are rate constants of the slow and fast kinetic phases, respectively, observed in the refolding experiments, and α_1 and α_2 are the corresponding relative amplitudes^{29,30}. The burst-phase amplitude refers to the amplitude of the optical signal that occurs within the dead-time of stopped-flow mixing (<4 ms) relative to the total amplitude of the signal that is observed. When folding is monitored by a change in ANS fluorescence, the entire increase in fluorescence occurs in the burst phase, but the decrease that follows occurs in two phases with rate constants $23 \pm 4 \text{ s}^{-1}$ and $7 \times 10^{-3} \pm 3 \times 10^{-3} \text{ s}^{-1}$ (Fig. 3a), corresponding in magnitude to λ_2 and λ_1 , respectively³⁰. Kinetic amplitudes in the single-jump experiments reported here do not reflect the relative populations of U_F and U_S , which can only be determined accurately by double-jump experiments²⁹. This is because fast folding reactions occur not only in the folding pathway of U_F but also in that of U_S ^{27,30}. Errors shown for the kinetic parameters represent the typical deviations seen in measurements from at least three different experiments in each case.

Formation of secondary structure was monitored by measurement of the change in ellipticity at 222 nm. There is no change in ellipticity during the burst phase, within the mixing dead time (<4 ms). Thus the burst-phase intermediate I_{M1} has no optically active secondary structure. The entire change in ellipticity at 222 nm is observed upon refolding in 1.0 M GdnHCl (Fig. 1), and occurs in two kinetic phases, as described in Table 1.

Formation of tertiary structure was monitored using two different probes. The change in ellipticity in the near-ultraviolet at 270 nm is due to the gain in asymmetric environments by the aromatic residues upon refolding, and is therefore a more sensitive probe for the formation of the final tertiary structure. Figure 2a shows the kinetics of the change in ellipticity at 270 nm on refolding. The entire change in ellipticity at 270 nm can once again be observed, and occurs in two kinetic phases (Table 1). Figure 2b shows the kinetics of the increase in intrinsic trp fluorescence, as described in Table 1. As reported previously^{27,29}, the entire increase in trp fluorescence is observed. The increase in intrinsic trp fluorescence that occurs upon refolding is due to the increase in the quantum yield of trp fluorescence as the tryptophans are sequestered from water as the structure forms.

The hydrophobic probe ANS has been shown to bind both burst-phase intermediates and later kinetic intermediates of the folding pathways of proteins^{7,16,25,26,30}. ANS binds to non-polar surfaces, presumably in the hydrophobic core, which in folding intermediates may still be solvent accessible. Such binding is therefore commonly used as a measure of the compactness of the bound intermediates^{6,22}. In the case of barstar, ANS binds to the burst-phase intermediate, I_{M1} ³⁰, which must therefore possess

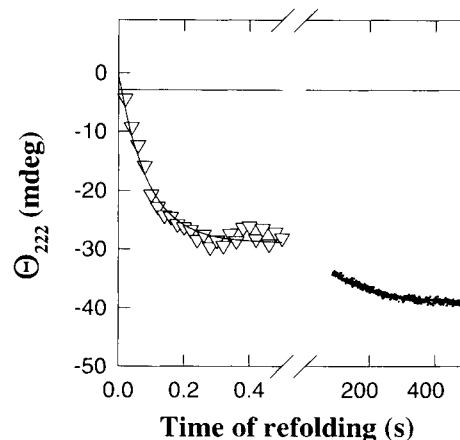
non-polar surfaces in a solvent-accessible hydrophobic core and is presumably compact. Figure 3a illustrates the kinetics of the change in ANS fluorescence during folding in 1.0 M GdnHCl. After the burst-phase increase in fluorescence on ANS binding to I_{M1} , it decreases in two phases (Table 1) as the hydrophobic core becomes inaccessible to the solvent.

To confirm that I_{M1} is indeed compact, the kinetics of the collapse of unfolded barstar molecules upon transfer to refolding conditions were determined by measurement of the kinetics of restoration of fluorescence energy transfer^{23,25} between the trp residues and fluorescent dansyl labels placed on the two cysteine sulphhydryls. Barstar has three trp residues distributed at both ends of its bundle-like structure, and thus fluorescence energy transfer averaged over all 6 trp-dansyl pairs can be used as a measure of overall compactness of the molecule⁶. On refolding, 70% of the energy transfer seen in fully folded protein is restored in a burst phase (4 ms) Fig. 3b. Because no increase in the intrinsic trp fluorescence quantum yield is observed in the burst phase Fig. 2b, this result suggests that unfolded barstar molecules collapse within 4 ms of transfer to refolding conditions. Complete energy transfer is restored in two observable kinetic phases (Table 1) which indicate either: further compaction of folding molecules; or an increase in the efficiency of energy transfer which would result from the increase in the quantum yield of trp fluorescence found to occur when the protein continues to fold after its initial collapse (Fig. 2b and Table 1).

Many proteins fold by means of early molten-globule intermediates, in which secondary structure appears to be stabilized in part by productive hydrophobic contacts^{4,5,15,26}. In the case of

FIG. 1 Formation of secondary structure. The change in ellipticity at 222 nm with time of refolding is shown. Barstar that had been unfolded to equilibrium in 6.0 M GdnHCl was refolded in 1.0 M GdnHCl at pH 7. Two kinetic phases are seen, and the results of a nonlinear least-squares fit of the data to the sum of two exponentials are shown in Table 1. The change in ellipticity upon refolding extrapolates back to the ellipticity of the protein in 6.0 M GdnHCl at time zero, within the experimental error in the determination of the unfolded protein baseline, represented by a horizontal solid line.

METHODS. The purification of barstar has been described previously²⁸. Stopped-flow circular dichroism studies of folding were performed using a Biologic SFM-3 stopped-flow machine interfaced to a Jasco J720 spectropolarimeter, using the optical and mechanical interface provided for the purpose by Jasco. A 1.5-mm pathlength FC15 cuvette was used for all circular dichroism studies. Barstar (20 μM) in unfolding buffer (6.0 M GdnHCl, 5 mM sodium phosphate, 250 μM EDTA and 300 μM DTT, pH 7) was mixed with 220 μL refolding buffer (0.55 M GdnHCl, 5 mM sodium phosphate, 250 μM EDTA and 300 μM DTT, pH 7), using a flow rate of 5 mL s^{-1} . The concentration of protein during folding was 30 μM . For data acquisition by the J720 at 222 nm, the response time was set to 16 ms. The data shown represents the average of 8 separate traces. The data were fitted using the nonlinear least-squares routine in the SigmaPlot for Windows graphics software to obtain the values of



λ_1 and λ_2 shown in Table 1. In this experiment, and those shown in Figs 2 and 3, the unfolded protein baseline was obtained by diluting the unfolded protein into unfolding buffer in a similar manner.

barstar, our data indicate that I_{S1} is such an intermediate. Its formation is accompanied by the fast change in intrinsic trp fluorescence^{27,30}, the rate of which is similar to that of the fast change in far-ultraviolet circular dichroism (Table 1). Thus I_{S1} possesses partial secondary structure as well as partial tertiary structure.

Our data show that a very early compact, structureless globule, I_{M1} , formed by a nonspecific hydrophobic collapse, precedes the molten, globule-like intermediate I_{S1} on the folding pathway. Previous studies³⁰ had suggested that the hydrophobic contacts in I_{M1} were nonspecific: the ANS binding ability of I_{M1} decreases linearly and not cooperatively with an increase in concentration of GdnHCl from 1 to 2 M. We have shown I_{M1} to have no optically active secondary structure. This result is surprising

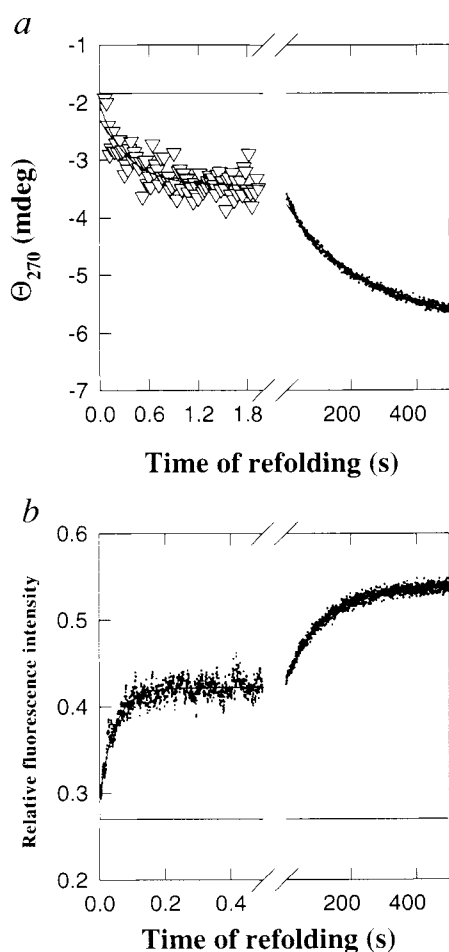


FIG. 2 Formation of tertiary structure. The changes in ellipticity at 270 nm (a) and fluorescence at 320 nm (b) with refolding are shown. Barstar that had been unfolded to equilibrium in 6.0 M GdnHCl was refolded in 1.0 M GdnHCl at pH 7. In both a and b, the horizontal solid line represents the unfolded protein baseline.

METHODS. a, Barstar (40 μ l) in unfolding buffer (6.0 M GdnHCl, 5 mM sodium phosphate, 250 μ M EDTA and 3 mM DTT, pH 7) was mixed with 200 μ l refolding buffer (5 mM sodium phosphate, 250 μ M EDTA, pH 7), using a flow rate of 4 ml s⁻¹. The concentration of protein during folding was 150 μ M, and that of DTT was 500 μ M. The data shown represent the average of 17 separately acquired traces. The values obtained for λ_1 and λ_2 by nonlinear least-squares analysis of the data are shown in Table 1. b, The measurement of the kinetics of folding by monitoring the change in intrinsic tryptophan fluorescence at 320 nm on excitation at 287 nm has been described previously^{29,30}. The concentration of protein during refolding was 5 μ M, and the buffers were the same as described for the far-ultraviolet data in Fig. 1. The values obtained for λ_1 and λ_2 by nonlinear least-squares analysis of the data are shown in Table 1.

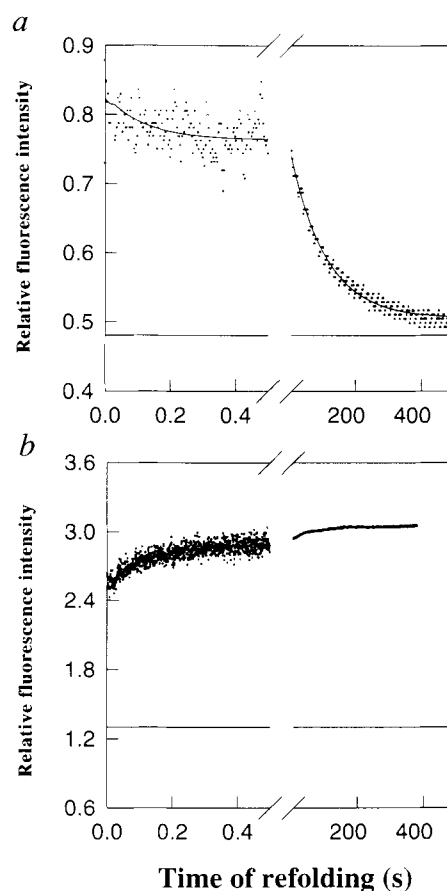


FIG. 3 Formation of a collapsed structureless globule in 1.0 M GdnHCl at pH 7. The change in ANS fluorescence at 460 nm (a) and restoration of trp-dansyl fluorescence energy transfer with refolding (b) are shown. In a, the ANS fluorescence for binding to fully unfolded protein in 6.0 M GdnHCl and fully folded protein in 1.0 M GdnHCl are also indicated by a horizontal solid line. In b, the fluorescence at 500 nm of the dansyl group bound to fully unfolded protein in 6.0 M GdnHCl is shown by a horizontal solid line to show that the efficiency of energy transfer in unfolded barstar is 40% of that in the fully folded protein.

METHODS. The use of ANS to monitor the folding kinetics of barstar has been described previously³⁰. The concentration of barstar during refolding was 10 μ M, and that of ANS 50 μ M. The folding kinetics of barstar, when monitored by the change in intrinsic tryptophan fluorescence, are identical in the presence and absence of ANS³⁰. This indicates that ANS does not perturb the folding pathway of barstar. For the fluorescence energy transfer experiments, 64 μ M barstar was reacted with 3 mM 1,5 IAEDANS (5-(((2-iodoacetyl)-N-methylamino)ethyl)amino)naphthalene-1-sulphonic acid) in 6.0 M GdnHCl, 20 mM sodium phosphate buffer at pH 7 for 1 h. Excess dye was then removed by gel filtration using a PD-10 column (Pharmacia). The extent of labelling was determined by measurement of the absorbance of the dye bound at 337 nm, and the absorbance of the protein at 280 nm (ref. 23); for the data shown in b, the extent of labelling was 1 dansyl per barstar molecule. Trp fluorescence was excited at 287 nm, and dansyl fluorescence was monitored using a 500 nm Oriol bandpass filter of bandwidth 50 nm. Labelled barstar (20 μ l) in unfolding buffer (6.0 M GdnHCl, 5 mM sodium phosphate, 250 μ M EDTA and 100 μ M DTT, pH 7) was mixed with 220 μ l refolding buffer (0.55 M GdnHCl, 5 mM sodium phosphate, 250 μ M EDTA, 100 μ M DTT, pH 7), using a flow rate of 4 ml s⁻¹. In a control experiment, the rates of folding of the labelled protein were shown to be similar to those of unlabelled protein, when monitored by intrinsic trp fluorescence (data not shown). Another control experiment showed that the activity of the labelled protein is 75% of that of unlabelled barstar (S. Ramachandran, personal communication). In a third control experiment (data not shown), results similar to those shown here were obtained both for the relative amplitude of the burst phase and the subsequent kinetics when the extent of labelling was increased to 1.6 dansyls per barstar molecule by changing the conditions for labelling.

because, for other proteins, burst-phase intermediates display secondary structure even when folding is performed at relatively higher concentrations of denaturant than those used here^{16,31}. It is unusual to find hydrophobic collapse preceding secondary structure formation for the folding of any protein in any folding condition. □

Received 31 May; accepted 4 September 1995

1. Ptitsyn, O. B. *Dokl. Akad. Nauk. SSSR* **210**, 1213–1215 (1973).
2. Kim, P. S. & Baldwin, R. L. A. *Rev. Biochem.* **51**, 459–489 (1982).
3. Udgaonkar, J. B. & Baldwin, R. L. *Nature* **335**, 694–699 (1988).
4. Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E. & Razgulyaev, D. I. *FEBS Lett.* **262**, 20–24 (1990).
5. Jennings, P. & Wright, P. E. *Science* **262**, 892–896 (1993).
6. Ptitsyn, O. B. in *Protein Folding* (ed. Creighton, T. E.) 243–300 (Freeman, New York, 1992).
7. Ptitsyn, O. B. *Protein Engng* **7**, 593–596 (1994).
8. Robson, B. & Pain, R. H. *J. molec. Biol.* **59**, 237–259 (1971).
9. Dill, K. A. *Biochemistry* **24**, 1501–1519 (1985).
10. Gutin, A. M., Abkevich, V. I. & Shakhnovich, E. I. *Biochemistry* **34**, 3066–3076 (1995).
11. Finkelstein, A. V. & Ptitsyn, O. B. *Prog. Biophys. molec. Biol.* **50**, 171–190 (1987).
12. Chan, H. S. & Dill, K. A. *J. chem. Phys.* **90**, 492–509 (1989).
13. Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S. & Nagamura, T. *FEBS Lett.* **221**, 115–118 (1987).
14. Kuwajima, K., Semisotnov, G. V., Finkelstein, A. V., Sugai, S. & Ptitsyn, O. B. *FEBS Lett.* **334**, 265–268 (1993).
15. Radford, S. E., Dobson, C. M. & Evans, P. A. *Nature* **358**, 302–307 (1992).
16. Mann, C. J. & Matthews, C. R. *Biochemistry* **32**, 5282–5290 (1993).
17. Matthews, C. R. A. *Rev. Biochem.* **62**, 653–683 (1993).
18. Roder, H. & Elove, G. A. in *Mechanisms of Protein Folding* (ed. Pain, R. H.) 26–54 (IRL, New York, 1994).
19. Khorasanizadeh, S., Peters, I. D., Butt, T. R. & Roder, H. *Biochemistry* **32**, 7054–7063 (1993).
20. Varley, P. et al. *Science* **260**, 1110–1113 (1993).
21. Gujjarro, J. I. et al. *Biochemistry* **34**, 2998–3008 (1995).
22. Feng, H.-P. & Widom, J. *Biochemistry* **33**, 13382–13390 (1994).
23. Kawata, Y. & Hamaguchi, K. *Biochemistry* **30**, 4367–4371 (1991).
24. Elöve, G. A., Chaffotte, A. F., Roder, H. & Goldberg, M. E. *Biochemistry* **31**, 6876–6883 (1992).
25. Semisotnov, G. V. et al. *Biopolymers* **31**, 119–128 (1991).
26. Itzhaki, L. S., Evans, P. A., Dobson, C. M. & Radford, S. E. *Biochemistry* **33**, 5212–5220 (1994).

27. Schreiber, G. & Fersht, A. R. *Biochemistry* **32**, 11195–11203 (1993).
28. Khurana, R. & Udgaonkar, J. B. *Biochemistry* **33**, 106–115 (1994).
29. Shastry, M. C. R., Agashe, V. R. & Udgaonkar, J. B. *Protein Sci.* **3**, 1409–1417 (1994).
30. Shastry, M. C. R. & Udgaonkar, J. B. *J. molec. Biol.* **247**, 1013–1027 (1995).
31. Sugawara, T., Kuwajima, K. & Sugai, S. *Biochemistry* **30**, 2698–2706 (1991).

ACKNOWLEDGEMENTS. We thank M. K. Mathew for discussion and S. Ramachandran for checking the activity of labelled protein. This work was funded by the Tata Institute of Fundamental Research and by the Department of Biotechnology, Government of India.

CORRECTION

Regulation of neural induction by the Chd and Bmp-4 antagonistic patterning signals in *Xenopus*

Yoshiki Sasai, Bin Lu, Herbert Steinbeisser & Eddy M. De Robertis

Nature **376**, 333–336 (1995)

THE PCR primer sequences for NF-M and NCAM in the legend of Fig. 1 were accidentally deleted. The primer sequences for NF-M are: F, 5'-GAACAGTACGCCAAGCTGACTG, and R, 5'-GCAGCAATTTCTATATCCAGAG (28 cycles); those for NCAM (accession no. M25696) are: F, 5'-GCGGGTACCTTCTAATAGTCAC, and R, 5'-GGCTTGGCTGTGGTTCTGAAGG (28 cycles). We thank R. Harland for pointing out our mistake. □

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

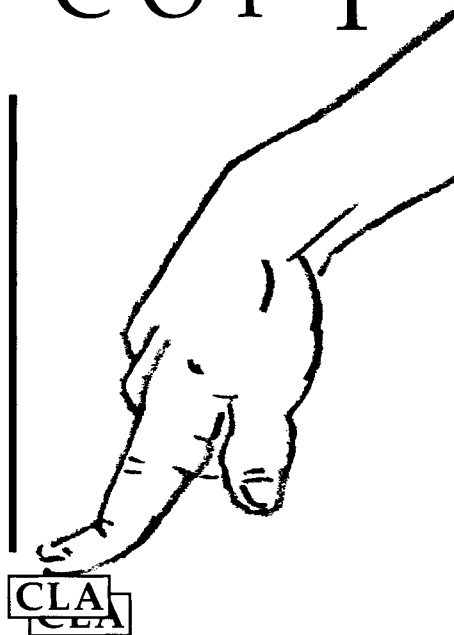
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised infringing photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931
Fax: 0171 436 3986

A new non-isotopic detection system for immunoassays

Daniel R. Deaver

A new technology for conducting immunoassays that uses an electrochemiluminescence detection system, eliminates the use of radioisotopes and offers improved assay performance.

The electrochemiluminescent (ECL) detection system is a new technology for quantification of the binding of any two molecules that join with specificity. Consequently, this system is well suited for conducting immunoassays. Three components are required to take advantage of this technology: (1) a compound labelled with a molecular species capable of generating chemiluminescence, (2) a detection buffer containing tripropylamine (TPA), and (3) the analyser, a proprietary electrochemiluminescence detection system, Origen, developed by Igen (ref. 1). In contrast to conventional chemiluminescence, with ECL the chemiluminescent species are generated at the surface of an electrode. Current applications utilize an *N*-hydroxysuccinimide ester (NHS) of a ruthenium chelate to label proteins and small haptens. The conjugation and separation procedures are simple. Stability of ruthenium-labelled material is excellent. A major advantage of this technology is the variety of ways that assays can be formatted for the instrument. The detection system allows investigators the flexibility of designing both separation and nonseparation assays.

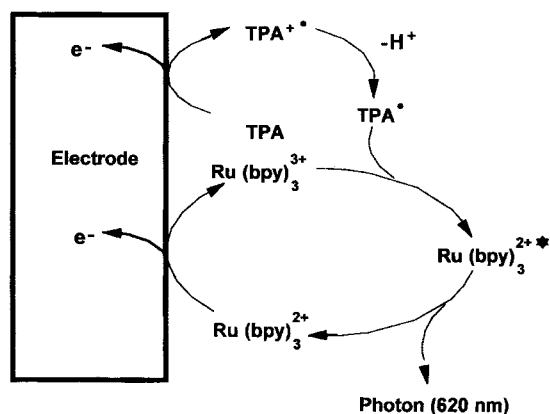


FIG. 1 Mechanism for generating the ECL signal.

When compared with radioimmunoassay (RIA) and enzyme-linked immunosorbent assay (ELISA) techniques, assays using this technology have a lower investment of labour, generally are more sensitive and have a greater dynamic range. For nonseparation assays, the capture, wash and detection steps are conducted automatically. Consequently, there are no plate coating/washing steps or decanting/aspiration procedures as typically required for ELISA and RIA techniques,

respectively. These features make the ECL detection system an attractive alternative to conventional RIA, ELISA and conventional chemiluminescence-based assays.

Conjugation procedures

The incorporation of a reactive species capable of generating chemiluminescence must be incorporated into the assay system. Currently, an *N*-hydroxysuccinimide ester of ruthenium (II) tris-bipyridine chelate (TAG-NHS ester) is available from Igen for labelling proteins and haptens. This material is supplied as solid, which is stable for at least 18 months when stored desiccated at -10 to -30 °C. We have successfully conjugated a variety of proteins to TAG-NHS in our laboratory, including polyclonal and monoclonal antibodies, luteinizing hormone (LH), insulin-like growth factor-1 (IGF-1) and salmon calcitonin. The *N*-hydroxysuccinimide ester of ruthenium combines with ϵ -amino group of lysine to form a stable amide bond. Immediately before use, the TAG-NHS ester is dissolved in anhydrous dimethylsulphoxide (DMSO). The amidation reaction is completed within 30–60 min, and is terminated

by the addition of 2 M glycine. With simple modifications the procedure can be adjusted to label as little as 10–25 μ g or as much as 1 mg of protein. Separation of TAG-conjugated compounds from unreacted TAG-NHS is relatively easy. For proteins greater than 20K we have found gel filtration using a 1×20 cm column of Sephadex G-25 to be a simple and effective method. For TAG-LH and TAG-IGF-1 the elution patterns from gel-filtration columns were found to be identical to that observed for 125 I-labelled

compounds. Low molecular weight peptides, such as salmon calcitonin, can be separated from unreacted TAG-NHS by using an Amicon Centricon-3 concentrator. Conjugated materials are diluted with an appropriate buffer containing 1–3 per cent bovine serum albumin and stored at -80 °C. We prepared our first TAG conjugates over one year ago. Aliquots of materials from these initial preparations are still being used with no signs of deterioration over time.

ECL analyser

The analyser is a compact unit measuring $41 \times 30 \times 30$ cm (L×W×H), which includes a personal computer and printer. The analyser integrates fluid-handling components (including an automatic sampler), luminometer, potentiostat, and electrochemical flow cell and a round carousel sample changer. The carousel holds fifty 12 mm $\times$ 75 mm polypropylene tubes. Although the capacity of the sample changer is adequate for most research settings,

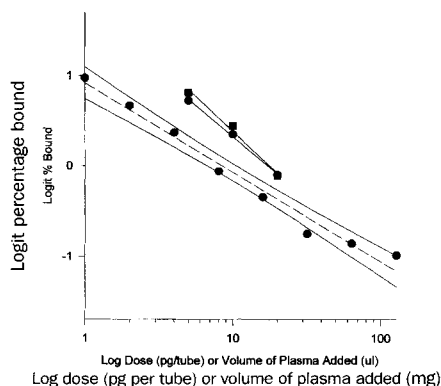


FIG. 2 Typical standard curve for the assay of salmon calcitonin and dose-response relationships with varying amounts of chicken plasma. Solid circles represent the means of duplicate data points; the dashed line, the calculated best-fit linear regression; and the solid black lines, the 95 per cent confidence intervals. Data for ECL measurements were converted to per cent bound, and the logit of per cent bound is plotted against log of mass of salmon calcitonin.

the current configuration of the unit is not well suited for large-scale clinical laboratories. In addition, some researchers may find the capacity of the carousel limiting, especially when compared with microtiter plate readers and large capacity gamma counters. Based on our experiences, one instrument can handle a throughput of 50,000 tubes per year. Also, significant reductions in labour requirements due to stability of reagents has allowed our laboratory to change our approach and to perform frequent small assays (200–300 tubes), instead of larger ones centred around iodination dates.

The major component of the system is the electrochemical flowcell, which contains the working electrode and counter electrodes for initiation of the ECL reaction. Details of this system have been published elsewhere^{1,2}. In

the normal sequence of sample analysis, a precise volume of sample is removed from the sample tube and delivered to the flowcell by a peristaltic pump. The carousel unit also serves as a vortex, and tubes are shaken before each sampling step to ensure good mixing of the sample. The mechanism for

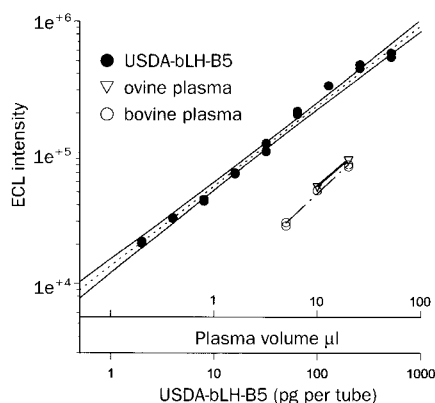


FIG. 3 Representative standard curve for the assay of LH using a sandwich-type format. Solid circles represent the mean of duplicate data points, the dashed line, the calculated best-fit linear regression and the solid black lines, the 95 per cent confidence intervals. Dose responses of varying volumes of bovine and ovine plasma were parallel to the standard curve. The $\log_{10}$ of zero-corrected ECL values are plotted against the $\log_{10}$ of mass of LH.

generating the chemiluminescence is depicted in Fig. 1. Briefly, ruthenium(bpy) $_3^{2+}$ and tripropylamine (TPA) are oxidized at the surface of the electrode³. In order to position the TAG compound of interest near the electrode, an antibody coupled to a magnetic bead is used. A magnet located below the electrode will attract the TAG immune complex, allowing the ruthenium to be oxidized. TAG-labelled materials not bound to magnetic bead immune complex are washed from the flow cell by a stream of assay buffer. The oxidized TPA will spontaneously form a strong reductant, which reacts with the oxidized form ruthenium, forming an excited state of the compound at the surface of the electrode. The excited form of the label decays to its ground state, emitting a photon with a wavelength of 620 nm. Since this is a regenerating process, each ECL-active label can emit many photons during each measurement cycle. Light emitted during this process is detected and quantified by a photomultiplier tube located above the working electrode. A high pH cleaning solution is then drawn into the flow cell to clean the surface of the electrode, followed by TPA-containing assay buffer to eliminate the cleaning solution and to prepare the cell for the next measurement cycle. This entire sequence of events takes less than one minute per sample.

General operational functions, such as instrument calibration and systems checks, priming, flush procedures and automatic stand-by are controlled from the PC. Assay

parameter data relating to excitation method, luminometer settings and sample handling can be customized for individual assays. These custom settings can be saved and selected from a menu for future use. There is no automated data reduction for assays included with the software. However, data files can be easily imported into popular spreadsheets for processing.

The user has easy access to the pump head and tubing associated with the fluid-handling system allowing for easy replacement/adjustment of tubing. Occasional blockage of the fluid-handling system has been one of the main problems that we have encountered. Blockages have been traced to fibrin clots in samples or foreign material found in polypropylene tubes. We have been able to essentially eliminate this problem by centrifuging samples with visible signs of fibrin clots before use and by careful selection of vendors for tubes with minimal contamination. A second difficulty that we have encountered is a gradual increase in ECL values associated with ECL reading of buffer blanks over time. The cause of this phenomenon is not completely understood. Weekly decontamination of the analyser with bleach minimizes this problem.

Assay formats

A major advantage of this ECL immunoassay technology is the flexibility investigators have in designing the format of their assays. Both competitive and sandwich-type assays can be used. In our laboratory we have established assays that use this detection system for LH, salmon calcitonin, bovine IgG, bovine growth hormone and IGF-1. We have used both competitive and sandwich assay formats. Two examples are given to illustrate different advantages gained when converting from RIA to this technology.

Stability of ^{125}I -labelled salmon calcitonin and cost of reagents were major problems with our RIA for this hormone. The ECL assay protocol is similar to other commercial assays for this peptide, with two exceptions. First, TAG-salmon calcitonin is used in place of ^{125}I -labelled salmon calcitonin. Second, sheep anti-rabbit Dynal 280 beads are used to capture the immune complex at the surface of the electrode. The detection system has allowed us to make several improvements in the assay for salmon calcitonin. A typical standard curve is shown in Fig. 2. As with traditional immunoassay systems, data must be plotted using a log-logit transformation to achieve a linear relationship between dose and ECL intensity. The sensitivity of the assay has been improved slightly relative to our RIA, and is approximately 0.75 pg per tube. In addition, the total incubation time for the assay has been reduced from 36 h (for the RIA) to 12 h. A major improvement has been in the stability of the assay. The slope, y -intercept, and R^2 have averaged -0.983 , 0.899 and 0.99 . Within- and between-assay coefficients of

variation (CV) have averaged less than 5 per cent. These data are based on a series of 10 assays conducted over a 5-month period using the same TAG-salmon calcitonin preparation. Finally, we have been able to demonstrate excellent recovery (>95 per cent) of added hormone and parallelism when using 5 to 20 μl of sample (without extraction and chromatography) when using plasma from sheep and chickens (Fig. 2).

For the assay of LH we developed a sandwich-type assay. This approach has improved sensitivity and substantially reduced the time required to complete the assay. A high-affinity monoclonal antibody against bovine LH (ref. 4) was conjugated to TAG-NHS. The TAG-anti-bLH monoclonal, standard (or sample) and polyclonal antisera are incubated overnight at room temperature. Using this approach the sensitivity of the assay, expressed as picograms per tube, was increased by 20- to 25-fold when compared to a typical LH RIA (Fig. 3). The correlation between this ECL-based assay and our standard RIA is high ($R^2 > 0.95$). In addition, the standard curve spans more than three log units (1 to 2,000 pg). As for salmon calcitonin, the performance of this assay has proven to be stable over time. Over the past 7 months slope, y -intercept and R^2 have averaged 0.846, 3.75 and 0.99 when we used a linear regression of log zero-corrected ECL versus log dose of LH. The within-assay CV has averaged 4.5 per cent and between-assay CV less than 8 per cent. Figure 4 illustrates an example of pulsatile LH secretion in a

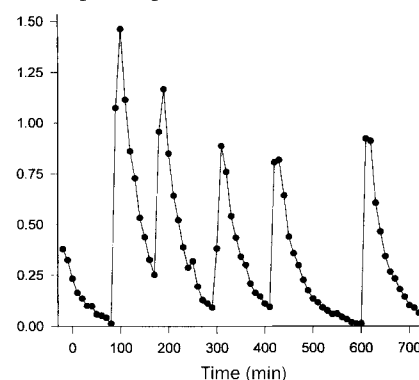


FIG. 4 Illustration of the pulsatile pattern of LH secretion in a juvenile male calf obtained using the ECL assay system. Note the number of data points that are below 0.25 ng ml^{-1} , which is the common limit of sensitivity reported for most LH radioimmunoassays using the same reference preparation.

young male calf during the juvenile stage of development using this assay system. Another interesting aspect of this assay format is our ability to measure LH from different species. The monoclonal antibody used is known to crossreact with LH from a wide range of species⁴. We measured rat LH using the TAG-monoclonal antibody against bovine LH in conjunction with standards and polyclonal anti-rat LH. With the improved

polyclonal anti-rat LH. With the improved sensitivity of the assay, only 20 µl of serum is needed per determination.

Comments

The ELC technology represents a novel approach for conducting immunoassays. Improvements over existing detection systems can include improved assay sensitivity and precision, reductions in time required to

complete assays, reduced labour needs, elimination of radioisotopes and excellent stability of TAG-labelled reagents. □

Daniel R. Deaver is a Professor of Reproductive Physiology in the Department of Dairy and Animal Science, and an Associate Director for the Center for Cell Research at The Pennsylvania State University, University Park, Pennsylvania,

16802-3503, USA. For more information, fill in reader service number 100.

1. Blackburn, G. F. *et al. Clin. Chem.* **37**, 1534-1539 (1991).
2. Yang, H. J., Leland, J. K., Yost, D. & Massey, R. J. *BioTechnology* **12**, 193-194 (1994).
3. Leland, J. K. & Powell, M. J. *J. electrochem. Soc.* **137**, 3127-3131 (1990).
4. Matteri, R. L., Roser, J. F., Baldwin, R. L., Lipovetsky, V. & Papkoff, H. *Domestic Anim. Endocr.* **4**, 157-165 (1987).

Immunological methods

New technology and techniques for understanding immunology are described below — including a serum protein quantification system, antibodies to proliferation markers, a human interleukin-10 ELISA and an unmasking reagent.

To help **differentiate mesotheliomas from carcinomas** in paraffin-embedded tissue sections, as well as in body-cavity fluids, BioGenex now offers antibody MOC-31 stain (*Reader Service No. 101*). The epithelial-cell-specific antigen (ESA), also known as the epithelial-specific cluster-2 antigen, is a 40K transmembrane glycoprotein of unknown function present on the membrane of epithelial cells. This monoclonal antibody, MOC-31, is the prototype reagent of cluster 2. The manufacturer states that MOC-31 is a highly reliable and sensitive marker used to help distinguish between epithelial and mesothelial cells. It reacts with virtually all normal epithelia and adenocarcinomas, but not mesotheliomas, and is useful for helping differentiate mesotheliomas from carcinomas. The product is currently available for research use only.

CPG's derivatized MPG (magnetic porous glass particles) is designed to provide rapid, reliable **magnetic separations for the isolation and purification of proteins, antibodies and oligonucleotides** (*Reader Service No. 102*). MPG is said to offer vast surface area (60 m² g⁻¹) for its three material types: uncoated, glyceryl and long-chain alkylamine particles. The uncoated particles provide free silanol groups to which ligands are first non-specifically adsorbed to the particle surface

and then stabilized by cross-linking with a homobifunctional reagent. In addition, uncoated particles can be easily modified to introduce almost any functional group with appropriate silanes. MPG glyceryl provides free hydroxyl groups that can be activated to covalently coupled ligands with primary or secondary amines. The non-ionic charge on the surface minimizes non-specific binding of the target biomolecule. With long-chain acrylamine particles, free amino groups provide attachment through a 15 Å spacer arm to which ligands can be covalently coupled.

Systems

A new system to **characterize molecular interactions**, called ConfoCor, has been jointly developed by Carl Zeiss and Evotec BioSystems (*Reader Service No. 103*). It is designed to examine molecular interactions in pharmaceutical research, molecular biology and immunology. The light of a laser is focused through confocal optics into a volume element of approximately 10⁻¹⁵ litres (about the volume of an *Escherichia coli* cell), and the molecules contained in this small volume are excited to fluoresce. The resulting fluorescent light is conducted through confocal optics to a single photon detector. If small particle numbers are contained in the volume element, the autocorrelation curve computed from the signal fluctuation will show larger amplitudes than in the case of higher concentrations. The amplitude of the autocorrelation function is a measure of the average number of particles contained in the volume under observation. The average time a particle spends inside the volume element and the characteristic particle diffusion coefficient can be calculated, if the volume is defined. This system is designed to measure nanomolar concentrations in

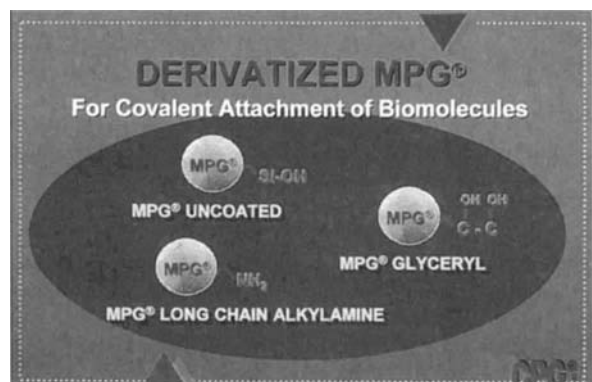


Molecular examination with ConfoCor.

times a short as a few seconds. The manufacturer states that ConfoCor is the first system to allow FCS routine examinations without requiring physical knowledge.

Signet Laboratories has introduced a series of new **alkaline phosphatase, anti-alkaline phosphatase (APAAP) detection systems for immunohistochemistry**, immunocytochemistry and immunoblot procedures (*Reader Service No. 104*). The APAAP detection system is designed with a pre-formed APAAP complex that should provide sensitivity without background staining caused by endogenous biotin or peroxidase molecules, particularly in cytology and freshly frozen tissue specimens. Potential research applications include the study of haematology specimens with preserved morphology, the study of fine-needle aspirates and cell smears, as well as staining of bone marrow specimens. The reagents are provided in a prediluted, easy-to-use format that is designed to combine sensitivity with consistent, reliable results. The system features colour-coded, ready-to-use reagents, a complete instruction manual and is available with choices of chromogens, including Fast Red, NBT/BCIP and new Fuchsin.

Incstar is now offering the SPQ II assay system for **serum protein quantification** — utilizing immunoturbidimetric methodology on automatic chemistry analysers, the system



CPG's magnetic porous glass particles.

can be used in testing panels for assessment of coronary artery disease, general immunology, nutritional studies, acute-phase reactants, and diabetes (*Reader Service No. 105*). The SPQ product line includes 15 analytes: α_1 -antitrypsin, apolipoproteins A-1 and B; fibrinogen, and lipoprotein(a); C3 and C4; CRP; haptoglobin; IgA; IgG; IgM; microalbumin; prealbumin; and transferrin. Each kit is sufficient for 100 to 300 determinations on commonly used centrifugal or discrete analysers. Calibration sets and antibody reagent sets are available separately, allowing adjustment to different usage rates. This system is available for research use only.

For immunophenotyping, absolute counting, reticulocyte enumeration and DNA analysis, Becton-Dickinson Immunocytometry Systems has announced the availability of the FACSCalibur automated **flow-cytometry system** (*Reader Service No. 106*). It is a four-colour benchtop system that combines both cell analysis and cell sorting applications. The system features touch-pad operation, an automated sample loader, object-oriented software, a broad range of reagents and an



A new serum-protein quantification system from Incstar.

Histosettes are available as either the regular or biopsy version, the latter having smaller holes to protect biopsy specimens. For clearing, the company offers xylene substitute, a greaseless, non-flammable and virtually odourless alternative to xylene, which utilizes a mixture of aliphatic hydrocarbons, the substitute does not irritate or sensitize normal skin. Shandon supplies a complementary xylene substitute mountant that is required for use with this clearing agent. Histoplast is a blend of paraffin waxes specially chosen to optimize processing, and it comes prefiltered to remove debris and in pellet form for rapid melting, the melting point is 56–57 °C.

Monoclonal antibodies

A wide range of **antibodies and related reagents for researching cell adhesion** are now available from TCS Biologicals and Upstate Biotechnology (*Reader Service No. 108*). These include monoclonal antibodies to the integrin family (chains b1, b5) including the laminin receptor (VLA-6, CDw49f), MAC1 (CD11b) and LFA-1 (CD11a); monoclonal antibodies to the immunoglobulin superfamily, including PECAM-1 (CD31), ICAM-1 (CD54) and NCAM (CD56). Also available are monoclonal antibodies to selectins, including PADGEM (GMP-140, P-selectin, CD62) and ELAM-1 (E-Selectin), along with a wide range of antibodies to other adhesion and platelet molecules. Of particular interest to researchers is the availability of monoclonal antibodies to human CD44 and specific epitopes expressed in alternatively spliced isoforms of CD44 (exons V4, V5, V7, V7-8 and V10).

Two new **monoclonal antibodies** have been released by Coulter — clones 49 IL-2R(p75)-FITC and CD38-RD1 are for use mainly in T-lymphocyte activation studies (*Reader Service No. 109*). The addition of these two new T-lymphocyte activation markers should

aid researchers to determine the potential clinical utility of activation markers. The first of the new activation markers is a monoclonal antibody to IL-2R (p75), which has been conjugated to FITC. IL-2R(p75), also known as CD122, is a subunit of the interleukin-2 (IL-2) receptor. It is normally expressed on activated T cells and natural killer cells, but is not expressed on resting T cells. It may be expressed on B cells and monocytes following activation. The second activation marker is a monoclonal antibody to CD38, which is conjugated to RD1. CD38 is a glycoprotein with 45K and 12K subunits, which appears on activated T- and B-lymphocytes and appear early in differentiation from stem cells.

Oncogene Science offers improved **antibody detection against the Ki-67 proliferation marker** (*Reader Service No. 110*). The MIB 1 monoclonal antibody specifically stains proliferating cell nuclei, but it is completely unreactive with cells in G0. Unlike other Ki-67 antibodies that only stain frozen sections, MIB 1 monoclonal antibody specifically stains frozen and paraffin-embedded tissue sections.

Sigma Biosciences now offers **monoclonal antibodies and conjugates to mouse leukocyte markers** — 17 monoclonal antibodies to mouse CD markers unconjugated and conjugated to FITC, PE and Quantum Red (*Reader Service No. 111*). These conjugated CD antibodies allow fluorescent labelling of cell-surface antigens and permit analysis and sorting of leukocyte subset populations. The mouse CD-specific antibodies are evaluated

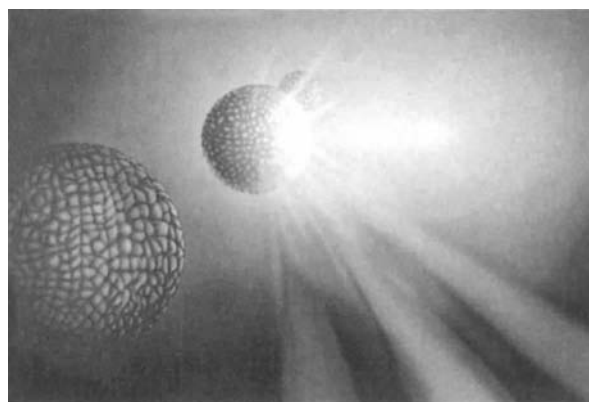


Becton-Dickinson's FACSCalibur system.

easy-to-use data management system. It was designed to support a wide range of applications for clinical and research settings including rare event analysis, dynamic immunophenotyping, immune function, platelet analysis, stem cell research, molecular flow techniques, and kinetics assays. Sorting occurs in a enclosed environment for enhanced safety when biohazardous samples are being processed. The dual laser technology is said to ensure high sensitivity, minimal compensation and maximum flexibility. In four-colour analysis, it can detect multiple fluorochromes and identify subset populations within a single sample.

Embedding and clearing

A full range of **accessories for embedding, clearing and mounting** is now available from Shandon (*Reader Service No. 107*). For processing and subsequent embedding, the company offers Super Mega cassettes — an extra large cassette for the processing of large blocks, whole organs or small animals.



Sigma Biosciences' antibodies to mouse CD markers.

for their performance in flow cytometry and formulated to obtain maximum per cent positive staining without significant elevation of background staining. Each vial contains 100 μ g of labelled purified rat immunoglobulins in approximately 0.4 ml (approximately 0.25 mg ml⁻¹). A product specific data sheet with protocols for flow cytometry applications is provided. One microgram of each conjugate is sufficient for staining 1×10^6 cells to maximum fluorescence intensity and per cent positive cells.

ELISAs

Autogen Bioclear has announced the availability of a new **ELISA for human interleukin-10**, the Innotech hIL-10 from Innogenetics (*Reader Service No. 112*). Interleukin-10 is a modulator of function in a variety of human lymphoid cells, a suppressor of macrophage function and a stimulator of B cells. IL-10 is capable of suppressing IFN- γ and numerous monokines in a variety of *in vitro* assay systems. The elevation of these same cytokines in anti-IL-10-treated mice suggests that these immunosuppressive properties will occur *in vivo* and that IL-10 may be a potent anti-inflammatory reagent in a variety of clinical situations. The ability of IL-10 to suppress a subset of T-cell-derived cytokines in addition to monokine suppression also suggests the potential for prolonging allograft survival and treatment for autoimmune diseases, for example, type 1 diabetes and multiple sclerosis. Each kit contains 2 $\times$ 96-well test plates in an 8-well strip format and all reagents and standards required for hIL-10 detection. Levels as low as 1.5 pg ml⁻¹ in serum and cell culture supernatant are detected in just over three hours.

The Quantikine **MCP-1 ELISA**, an immunoassay for the precise measurement of MCP-1 in human serum, plasma, culture, media and urine samples, is now available from R&D Systems (*Reader Service No. 113*). Monocyte chemoattractant protein 1 (MCP-1), also known as MCAF, LDCF or GDCF, is a member of the β or (C-C) subfamily of chemokines. The assay is designed to give



MCP-1 ELISA from R&D Systems.

consistent, reproducible results with a sensitivity of 5 pg ml⁻¹ across a range from 15.6 to 2,000 pg ml⁻¹. Normal blood level values for healthy individuals lie in the range 200–722 pg ml⁻¹. MCP-1 has, in the past, been measured using Boyden chamber or agarose gel chemotaxis assays. This is designed to be a rapid, specific and easy-to-use ELISA that offers an alternative for those investigating inflammation and inflammatory diseases.

Santa Cruz Biotechnology now offers **bcl 2 (100) and bcl 2 (4C11) mouse monoclonal IgG antibodies** (*Reader Service No. 114*). The bcl gene was isolated at the chromosomal breakpoint of t(14;18)-bearing follicular B-cell lymphomas. For researchers studying the complex patterns of signal transduction,

the new antibodies are said to be useful for western blotting, immunoprecipitation, ELISA and immune complex protein kinase assays, as well as other research applications.

Accessories

Jerini Bio Tools now offers **intelligent peptide libraries in immunology**, developed from different types of L-, D-, branched-, cyclic- and scanning-peptide libraries and synthesized covalently on continuous cellulose membranes using the Spot Synthesis technique (*Reader Service No. 115*). The manufacturer states that novel synthetic strategies combined with the use of robots allow economical, rapid synthesis of highly complex libraries generated by DIGEN software. Sequence diversity is generated for screening purposes, with any target molecule of interest, such as proteins, nucleic acids and metals. Selected peptides are further improved for binding by various library techniques to generate bioactive substances and analytical molecules. Peptide libraries of several billion compounds can be constructed and screened in as little as six weeks. Soluble peptide and non-peptide libraries will be announced shortly.

Accu-Tuf, a **universal antigen unmasking reagent** from Accurate Chemical, is designed not only to eliminate the need for multiple antigen recovery solutions but also to enhance the antigenicity of target epitopes in paraffin and frozen specimens (*Reader Service No. 116*). Used as an initial step in standardizing monoclonal and polyclonal antibody immunostaining procedures, the product is designed to significantly intensify specific stains while diminishing non-specific background stain. It is said to contain no heavy metals or other carcinogenic compounds and does not require a microwave or other special equipment.

For **immunostaining**, the new Vectastain Universal Quick kit contains a streptavidin-peroxidase preformed complex, blocking serum and a universal biotinylated antibody for use with any mouse, rabbit, rat and goat primary antibodies (*Reader Service No. 117*). The kit is said to utilize advances in biotin/streptavidin preformed complex technology to provide outstanding sensitivity with detection times as short as 15 min following incubation with primary antibody. Reagents are provided as stable stock solutions in convenient drop-bottle dispensers, and working solutions can be used immediately after dilution. Each kit provides sufficient reagents to stain approximately 500 frozen or paraffin-embedded tissue sections and is suitable for use with any automated immunostainers. □

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

OLIGONUCLEOTIDES

Synthesised, purified and shipped in 3 days

£2.50 per base

(All inclusive)

0.2 μ M scale (approx. 500 μ g)

- Antisense (completely or partially sulphurised)
- Extensive range of labelling and modifications
- Full design and analysis facilities
- Gene cassettes, bulk order and contract discounts

Molecular Medicine Unit

King's College School of Medicine & Dentistry
Tel: 071 346 3126 Fax: 071-733 3877

Reader Service No.10



Gramsch Laboratories

- New: δ - & μ - Opioid Receptor Antiserum
- G-Protein Antisera - α , β , γ -subunit
- G-Protein Immunoprecipitation Gels
- G-Protein Peptide Fragments
- Monoclonal Opioid-Antibody 3E7
- YGGF... - specific; receptor similarity; appl.: e.g. peptide com. library tests
- Cyclic Nucleotide Antisera
- Antibody/Peptide Production Service

Please, order today, we ship tomorrow
Phone 08138 1769 or 08131 5679-0
Fax 08138 1260 or 08131 5679-24

Gramsch GmbH

D-85247 SCHWABHAUSEN, Germany

Reader Service No.28

FLUORESCENT PROBES

DNA/RNA Stains *Fura-2 AM*
Ion Probes *1 mg DM 141.10*
pH Indicators
Enzyme Substrates *BCECF-AM*
Fluorescent Labels *1 mg DM 78.00*
Membrane Probes

EXCELLENT QUALITY - COMPETITIVE PRICES

LAMBDA FLUORESCENZTECHNOLOGIE GmbH

Grottenhofstr. 3, A-8053 Graz, AUSTRIA.

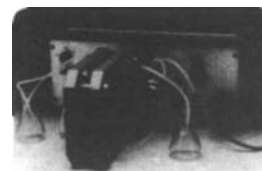
Tel.: ++43316/265352 Fax: ++43316/278586

In USA and CANADA call: 814-359-5064

Reader Service No.3

LIPOSOMES

REPRODUCIBLE - HOMOGENEOUS -
EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5-10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIANORM

P.O. Box 650126, D-81215 Munich
For USA call 1-800 Liposome

Reader Service No.25